2009

# EFFECT OF ANTIGENIC SITE MUTATIONS ON THE BINDING SPECIFICITY OF AN ANTI-HEMAGGLUTININ ANTIBODY TO H3N2 INFLUENZA VIRUS ISOLATES

Juliana Liambaya Hagembe
*Northern Michigan University*

# EFFECT OF ANTIGENIC SITE MUTATIONS ON THE BINDING SPECIFICITY OF AN ANTI-HEMAGGLUTININ ANTIBODY TO H3N2 INFLUENZA VIRUS ISOLATES

By

JULIANA LIAMBAYA HAGEMBE

THESIS

Submitted to
Northern Michigan University
In partial fulfillment of the requirements
For the degree of

MASTER OF SCIENCE

Graduate Studies Office

2009

SIGNATURE APPROVAL FORM

This thesis by Juliana Liambaya Hagembe is recommended for approval by the student's thesis committee in the Department of Chemistry and by the Dean of Graduate Studies.

Committee Chair: Dr. Mark D. Paulsen Date:

First Reader: Dr. Lesley Putman Date

Second Reader: Dr. Suzanne Williams Date

Department Head: Dr. Suzanne Williams Date

Dean of Graduate Studies: Dr. Cynthia Prosen Date

**OLSON LIBRARY**
**NORTHERN MICHIGAN UNIVERSITY**

# THESIS DATA FORM

In order to catalog your thesis properly and enter a record in the OCLC international bibliographic data base, Olson Library must have the following requested information to distinguish you from others with the same or similar names and to provide appropriate subject access for other researchers.

NAME: Hagembe, Juliana Liambaya

DATE OF BIRTH: December 15, 1981

ABSTRACT

EFFECT OF ANTIGENIC SITE MUTATIONS ON THE BINDING SPECIFICITY OF AN ANTI-HEMAGGLUTININ ANTIBODY TO H3N2 INFLUENZA VIRUS ISOLATES

By

Juliana Liambaya Hagembe

The hemagglutinin (HA) is the main antigenic glycoprotein found on the surface of the influenza virus and contains a sialic acid receptor-binding site that allows binding of the virus to a host cell. This important role makes it an ideal target for neutralizing antibodies. However, its ability to mutate and change every year at antigenic sites reduce or inhibit the binding of anti-hemagglutinin neutralizing antibodies, thereby allowing new subtypes to spread within non-immune species. Understanding the effect mutations of HA to antibody binding specificity, is not only fundamental to understanding the biological processes but also helpful in inhibitor design through rational design. Molecular Dynamics simulations and Molecular Mechanics-Poisson Boltzman Surface area (MM-PBSA) free energy calculations were used to study and compare the energetics of the binding of HAs from seven, Hong Kong, influenza, H3N2 Isolates from 1968, 1969, 1974, 1975, 1985, 1992, 1999 to the A/AICHI/68 (H3N2) anti-HA antibody. This study provides a bit more insight on the trend of mutations of HA and the route of evolution of influenza viruses. Results show an overall decrease in binding energy with increase in mutations and year progression. In addition, the association of HA to the antibody is driven by highly favorable van der Waals interactions whereas solute and solvent electrostatic energies contributed unequally to binding.

# DEDICATION

This thesis is dedicated to my brother, Joel Hagembe, and to my parents,
Drs. Peter and Bilha Hagembe

# ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor, Dr. Mark Paulsen, for his for his patience and guidance. He was always accessible and willing to help; as a result working on my thesis was a truly enjoyable and rewarding experience.

Besides my advisor, I would also like to thank Dr. Suzanne Williams and Dr. Leslie Putman for being part of my thesis committee.

I am also grateful to Dr. Paul Duby for offering me a Graduate Administrative assistantship with the Office of Institutional Research and Dr. Suzanne Williams for offering me an Adjunct Instructor position in the Chemistry Department; both of which allowed me to support myself through my masters.

In addition, my deepest gratitude goes to Professor Sandra Poindextor for her encouragement, wisdom and support and Mr. Arno wa Kangeri for his for assistance with refining the finished document.

Lastly and most importantly, I offer my regards and blessings to my Dad and Mom for both the monetary and emotional support they provided throughout my masters and my life in general; my siblings, because they are just the best; and all of those who supported me in any respect during the completion of the project.

This thesis follows the format prescribed by the American Chemical Society http://pubs.acs.org and the NMU Department of Chemistry.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

# LIST OF SYMBOLS AND ABBREVIATIONS

Alanine (Ala)…A
Arginine (Arg)…R
Asparagine (Asn) …N
Aspartic Acid (Asp)…D
Cysteine (Cys)……C
Glutamic Acid (Glu)……E
Glutamine (Gln)…Q
Glycine (Gly)…G
Histidine (His)…H
Isoleucine (Ile)…I
Leucine (Leu)…L
Lysine (Lys)…K
Methionine (Met)…M
Phenylalanine (Phe)…F
Proline (Pro)…P
Serine (Ser)…S
Threonine (Thr)…T
Tryptophan (Trp)…W
Tyrosine (Tyr)…Y
Valine (Val) …V
Alpha…α
Angstroms…Å
Atmosphere. …atm
Antibody…AB
Assisted Model Building and Energy Refinement…AMBER
Basic Load Alignment Search Tool…. BLAST
Central Processing Unit…CPU
Constant number of particles, temperature and pressure constant…NPT
Constant number of particles, temperature and volume constant…NVT
Emperical Conformational Energy program for Peptide Potentials …ECEPP
Free Energy Pertubation…FEP
Generalized Born…GB
Gibbs function…G
Helmholtz function…A
Hemagglutinin…HA
Hemagglutinin subtypes 1-15…H1- H15
Isolelectric… pI
kDA…kiloDaltons

Link, Edit and Parm .....LEaP
Linear Combinations of Pairwise Overlaps.....LCPO
Linear Interaction Energy.....LIE
Matrix protein.....M
Matrix protein 1.....M1
Molecular Dynamic.....MD
Molecular Mechanics.....MM
Neuraminidase.....NA or N
Neuraminidase subtypes 1- 9.....N1- N9
Non-structural proteins.....NS
Nucleocapsid protein.....NP
Picosecond.....ps
Poisson-Boltzmann.....PB
Polymerase.....P
Ribonucleic acid.....RNA
Ribonucleoprotein.....RNP
Simulated annealing with NMR-derived energy restraints.....Sander
Solvent -Accessible Surface Area.....SASA
Surface Area.....SA
Root Mean Square Deviation.....RMSD

# INTRODUCTION

Interactions between proteins play an essential role in many biological processes such as enzyme regulation, signal transduction and immune response. The binding of two structurally different proteins to a similar surface has been described on several occasions. Some of the interactions studied concern complexes between antigen and antibody where one of the wild-type molecular partners is able to bind to structural variants of the other: in most cases, these wild-type counterparts are formed by proteins with only a few amino acids difference [1-3]. Understanding how protein complexes form and what determines their specificity, binding constants and substitutions is not only fundamental to biological interpretation of the processes but also helpful in inhibitor design through a rational approach [4, 5]. The influenza virus, which creates multiple problems due to its ability to mutate and change every year, is a prime candidate for such a study.

## 1.1 Influenza Virus

### 1.1 1 Influenza virus overview

The influenza virus is a respiratory pathogen that causes influenza, an acute respiratory disease with prominent systemic symptoms. A wide variety of warm-blooded animals including birds, various wild and domesticated mammals and humans are vulnerable to infection by the virus. In humans, in addition to the symptoms directly attributable to infection by the influenza virus, the patient will also be at risk for complications such as the onset of pneumonia [6]. This is a particularly serious complication for the elderly and others with immuno-compromised systems. Today, pneumonia and influenza are listed as leading causes of death in the United States, responsible for 50,000 to 70,000 deaths annually [7]. The seasonal nature of influenza, the ease of transmission, the high rates of fatalities that it causes among the elderly and among patients under 5 years of age, have made influenza a key public health concern [8].

The *Orthomyxovirus* genus, to which influenza belongs, consists of three species: A, B, and C, each distinguished by antigenic differences in two of their stable internal proteins, Nucleoprotein (NP) and Matrix protein (M) [9]. In addition, the viruses differ in their pathogenicity and genome organization. Type A viruses are found in a wide variety of warm blooded animals including birds and humans and causes periodic worldwide epidemics (pandemics) in their host organisms, whereas types B and C are chiefly human pathogens and rarely cause pandemics [10]. The viruses of this genus are most commonly spherical (filamentous forms may also occur), approximately 80 nm to 120 nm in diameter, and have an eight-fold segmented, single-stranded RNA genome with negative

polarity containing between 890 and 2341 nucleotides each. An excellent review on the structure and replication strategy of influenza viruses has been published recently [11].

The virus has a core ribo-nucleoprotein that contains the genetic information of the virus wrapped up in protein. The eight separate viral RNA segments of the antisense RNA of both the influenza A and B virus (Figure 1), each code for a functionally important protein. The virus even with a fairly small genome is able to encode for a variety of proteins using different reading frames and alternate RNA splicing. The proteins encoded by the RNA segments include; Polymerase B2 protein (PB2), Polymerase B1 (PB1), Polymerase A (PA), Hemagglutinin (HA), Nucleocapsid protein (NP), Neuraminidase (NA or N) and Matrix protein (M). Each RNA segment is encapsulated by the NP to form the ribo-nucleoprotein (RNP) complex.

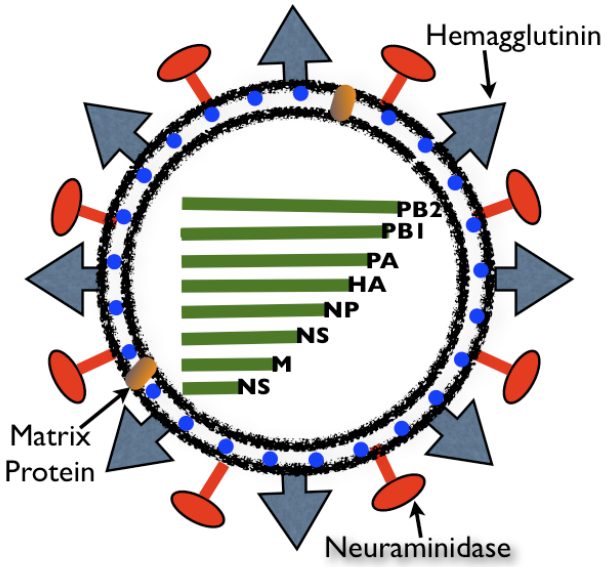


**Figure 1:** A diagrammatic representation of influenza A and B virus showing protein and RNA composition (Polymerase B2 (PB2), Polymerase B1 (PB1), Polymerase A (PA), Hemagglutinin (HA), Nucleocapsid (NP), Neuraminidase (NA or N), Non structural proteins (NS), Hemagglutinin (HA), and Matrix protein (M)).

The active RNA polymerase, which is responsible for replication and transcription, is formed from PB2, PB1 and PA. In addition to the polymerase activity, it has an endonuclease activity and is linked to the ribo-nucleoprotein. The NS1 and NS2 proteins have a regulatory function to promote the synthesis of viral components in the infected cell. The NP encapsulates each RNA segment thus forming a nucleocapsid. The nucleocapsid is surrounded by a shell of two Matrix proteins (M1 and M2). M1 constructs the matrix; and in influenza A viruses, M2 acts as an ion channel pump to lower or maintain the pH of the endosome of the infected cell. Overlying the matrix is the viral envelope consisting of a lipid bilayer, which has multiple copies of glycoproteins radiating from its surface. The characteristic rod-shaped "spikes" are HA and the interspersed mushroom-shaped clusters are NA. These HA and NA molecules are thought to pass through the envelope and interact with the underlying matrix protein.

HA and NA are the major antigens in influenza. HA mediates the first stages of virus infection; sialic acid binding and virus-cell membrane fusion [12], whereas NA is concerned with the release of progeny virions from the cell surface [13]. NA acts as an enzyme, cleaving sialic acid from the HA molecule, from other NA molecules and from glycoproteins and glycolipids at the cell surface [14]. The HA and the NA are used to further subdivide the influenza A virus into additional subtypes [15]. Currently 15 subtypes of HA (HI-H15) and nine subtypes of NA (N1-N9) have been discovered. The influenza virus continuously undergoes antigenic drift and antigenic shift to escape the host's acquired immunity. The two glycoproteins are constantly subjected to selection pressure by the host's defense mechanism. Antigenic drift, the accumulation of mutations in all influenza gene segments, is particularly prevalent in the surface glycoproteins (HA and

NA) [16-21]. The main cause of point mutations is thought to be due to a lack of proofreading ability by the RNA-dependent polymerase complex [22]. Antigenic shift, on the other hand, occurs as a result of genetic re-assortment of the genome segments from different influenza viruses. Antigenic shift causes the extinction of the current strain of the influenza virus due to replacement of the strain specific glycoprotein's (HA or NA) with new ones [23]. The source and host of such new viruses is thought to be animals such as birds or pigs.

To adequately describe a particular influenza virus isolate, the influenza virus type, the host species (omitted in the case of human origin), the geographical site of first isolation, serial number and year of isolation must be mentioned. In the case of influenza virus type A, the nomenclature also includes HA and NA subtypes in brackets. One of the parental avian strains of the current outbreaks of H5N1 of Asian lineage was isolated from a goose in the Chinese province, Guangdong: accordingly, it is designated A/goose/Guangdong/1/96 (H5N1) (Xu 1999), while the isolate originating from the human case of Asian lineage H3N2 infection from Aichi (1968) is referred to as A/Aichi/68 (H3N2).

So far only three HA subtypes and NAs from two subtypes have caused human pandemics: H1N1 in 1918, H2N2 in 1957 and H3N2 in 1968. Three small outbreaks have also arisen from avian subtypes (H5, H7, and H9). The avian subtypes managed to make a direct leap to humans from birds but their low transmissibility prevented major epidemics [24]. The next expected pandemic threat at the moment is from the H5N1 (bird) or H1N1 (swine) virus subtype [25].

Vaccination is the primary method for preventing influenza and its implications. So far inactivated-virus vaccines have provided essential protection when the vaccine antigens and the circulating viruses share high degree of similarity in the HA protein. Since new influenza virus antigenic variants emerge frequently from accumulation of point mutation in the HA protein (i.e. antigenic drift), influenza vaccines need to be updated frequently based on the results of global influenza surveillance; including clinical, virological and immunological surveillance. In virological surveillance, influenza viruses are characterized antigenically on the basis of the ferret serum antibody cross- reactivity. Antigenic variants selected serologically are then tested for antibody cross- reactivity in human sera to evaluate the potential cross-protection against the antigenic variant provided by the current vaccines and to select vaccine strains for the next season [7, 26].

### 1.1.2 Hemagglutinin overview

HA is the main influenza antigenic glycoprotein found on the surface of the influenza viruses (as well as many other bacteria and viruses) and plays an important role in virus life cycle. The term "hemagglutinin" arises from the protein's ability to cause red blood cells (erythrocytes) to clump together ("agglutinate") *in vitro*. The protein contains a sialic acid (N-acetylneuraminic acid) binding site, which allows binding of the virus to the cell that is being infected. It also induces membrane fusion allowing release of viral RNPs in to the cellular cytoplasm [1, 27]. The body of the HA molecule contains the stalk region and the fusiogenic domain that allows membrane fusion needed for the virus to enter a new cell. It consists of 562 - 566 amino acids, with a molecular weight of

approximately 220 kDa and has either two or three glycosylation sites [28, 29]. It spans the lipid membrane so that the major part, which contains at least 5 antigenic domains, is presented at the outer surface [30]. These antigenic sites, which surround the host receptor-binding site, are presented at the head of the molecule, while the feet are embedded in the lipid layer. Prominent mutations in the antigenic sites reduce or inhibit the binding of neutralizing antibodies, thereby allowing a new subtype to spread within a non-immune species [22, 31].

### 1.1.3 Hemagglutinin Structure

The atomic structure of HA was first published in 1981 by Wilson, Wiley and Skehel to a resolution of 3 Angstroms [1]. This was the first look at the structure of a viral membrane protein. They found that HA is an elongated homotrimeric [27] integral glycoprotein that is shaped like a cylinder and measures 135 Å from insertion in the envelope membrane to its tip.

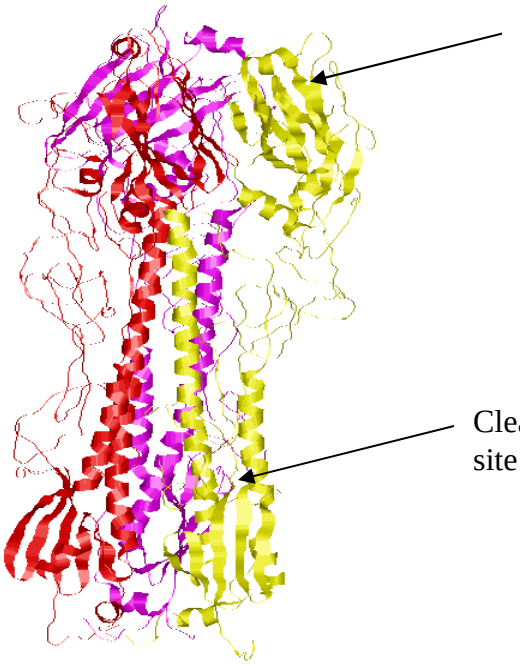


**Figure 2**: Ribbon representation of HA0 trimmer influenza virus. Each monomer (colored differently) possesses two important sites: **1)** The 'Receptor binding site' for virus attachment to the host lung epithelial cells via sialic acid containing host cell receptors. **2)** The 'cleavage site' where for full infectivity, the single chain (HA0) is cut into two chains (HA1 & HA2). At the N-terminal end of the HA2 chain is the fusion peptide, which is critical for subsequent membrane fusion events that lead to infection. *(Image created and modified using RasMol [PDB 1HGF][32]).*

The trimmer contains two distinct regions; a stem consisting of a triple stranded coiled-coil of alpha helices extending 76 Å from the membrane and a globular region of anti-parallel beta-sheet on top. The conserved host-receptor binding site was found to be located on the top globular portion. It is the amino acids that surround this region that have undergone change in most of the antigenic variants appearing in influenza virus epidemics. Each of the three monomeric subunits is synthesized as a precursor that is glycosylated and then cleaved into two smaller polypeptides (the $HA_1$ and $HA_2$ subunits). The six resulting chains are three $HA_1$s and three $HA_2$s.

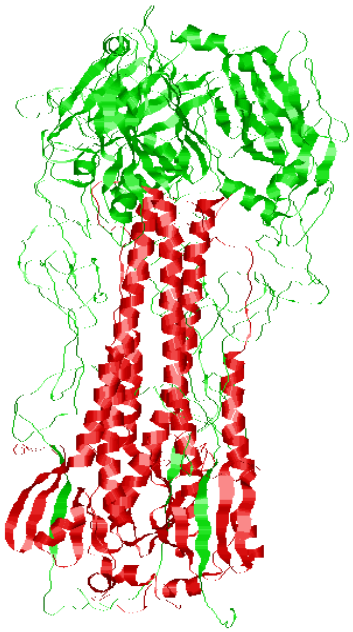

**Figure 3:** Ribbon diagram representation of HA showing the chains that result in the event of proteolytic cleavage of a trimmeric precursor protein. Three green $HA_1$ and three red $HA_2$'s *(Image created and modified using RasMol [PDB 1HGF]).*

The $HA_2$ is a helical chain anchored in the membrane and is topped by a large $HA_1$ globule. The $HA_1$ subunit (328 residues) is an elongate structure reaching from the

N-terminus at the viral membrane end of the molecule along the stem of the subunit before forming a globular tip. It then returns part way down the stem of the monomer, ending at the C-terminus part of its distal end. The $HA_1$ subunit forms an 8-stranded anti-parallel beta-sheet motif known as a "jelly roll."

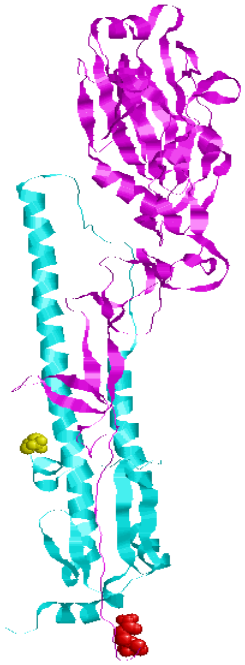

**Figure 4:** HA monomer, showing the $HA_1$ subunit (328 residues), an elongate structure reaching from the red N terminus at the viral membrane end of the molecule along the stem of the subunit before forming a globular tip and then returning part way down the stem of the monomer, ending at the yellow terminus tip. The long alpha-helix of H2 can also be seen (cyan**).** *(Image created and modified using RasMol [PDB 1HGF]).*

The loop that separates the 3rd and 4th strand of the jellyroll contains a short α-helix. Residues from one side of this α-helix and from residues near the top of the jellyroll form a pocket that is the sialic acid binding site for each monomer of the hemagglutinin trimmer. The binding sites contain the conserved amino acids - Y98, W153, H183 and Y195 – that form the base of the site which is limited at the top by a short α-helix (the 190-helix, residues 190–198), a loop-like structure at the front edge (the

130-loop, residues 133–138) and another loop that forms the left edge, near the intersubunit interface (the 220-loop, residues 220–229) [33]. Antigenic sites, which surround the host receptor-binding site, are presented at the head of the molecule. The receptor-binding sites in all HAs are at the membrane distal tip of each subunit in the trimer. Antigenic sites (antibody binding sites) residues in the HA1 chain, previously described by Brownlee and Fodor [34] are grouped into 5 regions: Ca1 (169 to 173, 208 to 296, 238 to 240), Ca2 (140 to 145, 224 to 226), Cb (78 to 83), Sa (128 to 129, 156 to 160, 162 to 167), and Sb (187 to 198)

'Fixed' changes, i.e. changes retained in HAs of viruses isolated in subsequent years, involved residues on the surface of HA, whereas about two-thirds of those not retained were found to be buried. These 'fixed' substitutions are thought to have been selected because they prevent antibody binding. This hypothesis is supported by the coincidence of the locations of these substitutions with the locations of amino acid substitutions detected in antigenic variant HAs that were selected by growing virus in the presence of monoclonal anti-HA antibodies [12, 35].

A notable feature of the $HA_2$ subunit (221 residues) is the two anti-parallel α -helices that form part of the stem of the molecule. One of the helices is among the longest known in globular proteins (75Å). The $HA_1$ and $HA_2$ chains of each monomer are connected via a single disulfide bond. The $HA_2$ subunit terminates in an α-helical structure near the protease cleavage site. The hemagglutinin trimmer is stabilized from interactions between the major $HA_2$ α -helices in the formation of a triple-stranded coiled coil in the trimmer's interior. The N-terminal (top) half of the coiled-coil super helix is tightly packed with several nonpolar residues in van der Waals contact around the three-

fold axis. The C-terminus end of the super helix expands away from the axis with polar and charged residues from each monomer experiencing electrostatic repulsion from like residues in the other monomers

The influenza virus binds to the cell surface by fixing the outer top of the HA to the sialic acid of a cell's glycoproteins and glycolipids. The sialic acid linkage to the penultimate galactose, either α 2→3 (in birds) or α 2→6 (in humans), determines host specificity [33, 36-38]. Avian influenza viruses generally show the highest affinities for α 2→3 linked sialic acid. In birds targeted by these viruses, α 2→3 linked sialic acid is the dominating receptor type in epithelial tissues of endodermic origin (gut, lung). In contrast, human-adapted influenza viruses, primarily access α 2→6 linked residues which predominate on non-ciliated epithelial cells of the human airway. The different receptor preferences help in preventing hassle-free transmission of avian viruses to humans thus creating a species barrier [39]. However, recently, it was shown that there is a population of ciliated epithelial cells in the human trachea, which also carry avian receptor-like glyco-conjugates at lower densities [40, 41]. Also chicken cells carry human-type sialic receptors at low concentrations. In pigs and quails, both receptor types are present at higher densities, thus making these species great hosts for mixing avian and human strains [42].

After attachment, the cell via a clathrin-coated pit receptor-mediated endocytosis process, as shown in Figure 6, takes up the virus. The time from entry to production of new virus is on average six hours [12]

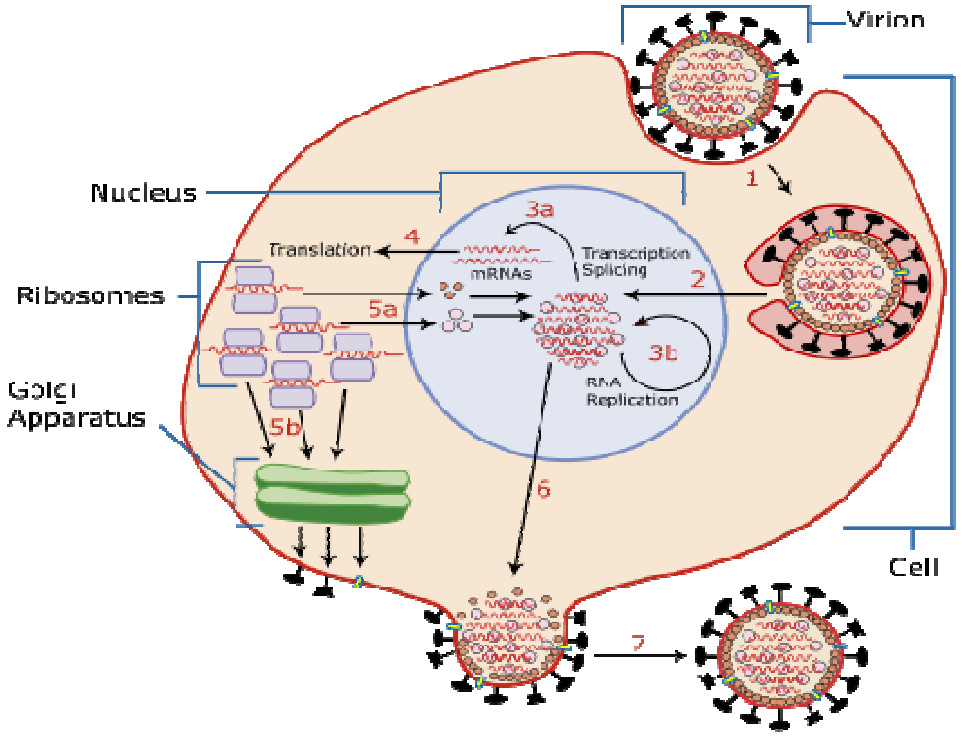


**Figure 6:** Replication cycle of influenza A virus. Binding and entry of the virus, fusion with endosomal membrane and release of viral RNA, replication within the nucleus, synthesis of structural and enveloped proteins, budding and release of virions capable of infecting neighboring epithelial cell [43].

### 1.1.4 Neutralization of Infectivity by Antibodies

Since viral attachment to cells is the first step in the infectious cycle, its inhibition would appear to be an effective way of preventing infection. HA contains the receptor binding site and studies have shown that anti-hemagglutinin antibodies neutralize virus infectivity in vitro, in addition to providing protection against infection [44]. A direct correlation between inhibition of virus binding to cells and neutralization of infectivity has been shown [45]. There are however, uncertainties about the mechanism of neutralization of influenza virus. Knoswo et al determined the structures of HA-AB

complexes with three distinct antibodies and was able to provide more insight on the mechanism by which anti HA antibodies neutralize infectivity [44]. The epitopes recognized by three antibody studies were located on the receptor-binding domain. Two of them overlapped with the receptor-binding domain and blocked access to it while the third had the ability to prevent the structural transition of HA that is required for fusion of virus and cellular membranes [46-48]. These three antibodies are thought to be representative of the range of neutralizing antibodies that react with HA.

The fact that the affinities of antibodies for HA are much stronger than the affinities of the receptor-binding site for sialic receptor analogues, allows antibodies to effectively block receptor binding. Comparison of the sizes of the receptor-binding site (800 $Å^2$) [33] , and of antibody–antigen interfaces of the monoclonal antibodies that contact amino acid residues that are components of the receptor-binding site, indicates that 'antibody footprints' (1200 and 1500 $Å^2$ ) [20, 49] are larger. Therefore, in addition to covering the receptor-binding site, the antibody would also cover HA residues that are not involved in binding to receptor. Mutations at these positions are what allow virus to escape neutralization without imposing selective pressure on residues of the receptor-binding site that could compromise its activity. Support for these conclusions is provided by the findings that the receptor-binding site is formed by conserved residues and that all the mutations which allow the virus to escape neutralization by antibodies that partially overlap the receptor-binding site, are at positions outside the site [48].

The relationship between antigenicity and immunogenicity remains unclear as our knowledge of the determinants of HA immunogenicity and of other factors that lead to the induction of antibodies with varying ranges of specificity in different infected

members of the population, is incomplete. It is nevertheless essential for an understanding of the pathway of antigenic drift and possibly also for attempts at effective vaccination of all sections of the population. Among the initial steps towards understanding antigenicity, is the comparison of antibody binding specificity towards different mutated HAs. The purpose of this study was to collect information regarding the change in binding specificity of HA to a specific antibody due to antigenic mutation of HA. Such information might enable prediction of subsequent structural and antigenic changes, as no identification of mutation preference or the ability to predict changes has been achieved. As a complementary approach to *in vitro* experiments (e.g. isothermal titration chemistry), molecular dynamics (MD) simulations could provide such information.

## 1.2 Computer Molecular Modeling

### 1.2.1 Molecular Modeling

Molecular modeling is a multidisciplinary field that encompasses laws and theories that stem from mathematics, physics, chemistry and biology and employs algorithms from computer science and information theory. It can be defined as the art and science of studying molecular structure and function through theoretical and computational techniques by building models or mimicking the behavior of molecules.

The role of computation in biology and biological chemistry has shown a steady increase over the past decades due to continued growth of computing power, in particular in the context of personal computers. Simulations that were difficult for yesterday's super computers can be carried out today using standard office workstations. Use of parallel

computers, which couple processors together in such a way that calculation is divided into small pieces with results being combined at the end, has also enhanced the ability to analyze, compare, and characterize large and complex data sets that are obtained from experiments on bimolecular systems.

In addition, the ability of computer simulations to complement experiments by providing averages and confirming distributions or interactions between parts of the system for a variety of properties of bimolecular systems that cannot be measured by experimental means, have rendered the method invaluable.

As a counterpart to experiments, Molecular Dynamic (MD) simulations are used to estimate equilibrium constants and dynamic properties of complex systems that cannot be calculated analytically. Modeling approaches thus symbolize the exciting interface between theory and experiment by filling in the many gaps, thus enabling building of better models and theories that ultimately make (testable) predictions [50, 51].

Molecular modeling studies usually involve three stages. In the initial stage, a model is selected to describe the intra and inter-molecular interactions in the system. The two most commonly used models are quantum mechanics and molecular mechanics; both enable the energy of any arrangement of atoms and molecules in the system to be calculated. They also allow the determination of energy variation in the system as positions of atoms and molecules change. The second stage is the calculation itself. Calculations such as energy minimizations, molecular dynamics and Monte Carlo simulation/conformation can be made. The last stage involves analysis of the calculation; to not only calculate properties, but to check that it was performed properly.

### 1.2.2 Biomolecular Force Fields (Molecular Mechanics)

The potential energy of intramolecular and intermolecular interactions is one of the most important variables in computer modeling. Molecular mechanics is based upon a rather simple model of these interactions within a system in addition to contributions from stretching of bonds, the opening and closing of angles and rotations about single bonds.

The basic concept is that, given a molecular structure (i.e., cartesian or internal coordinates for all the atoms), we can calculate the energy of interaction between all atoms. In this way, the energy values of the molecular conformations found in nature can be calculated. The intermolecular energy between two molecules interacting such as a drug molecule interacting with proteins or two proteins interacting with each other can also be calculated. The molecular potential energy is presumed to be a sum of energy terms that correspond to physical effects such as electrostatic interactions, dispersion and repulsion energies, bond distance and bond angle distortion. The energy is assumed to be a function of the number and type of chemical species within the molecule and the distance between all pairs of atoms (see equations below).

$$E = E_{bonds} + E_{angle} + E_{dihedral} + E_{non-bonded}$$

$$E_{non-bonded} = E_{electrostatic} + E_{van\ der\ Waals}$$

The set of parameters consisting of equilibrium bond lengths, bond angles, partial charge values, force constants and van der Waals parameters are collectively known as force fields and can be better described as the functional form and parameter sets used to denote the potential energy of a system of particles. Different implementations of

molecular mechanics use slightly different mathematical expressions, thus different constants for the potential functions. It should be noted that transferability is a key attribute of a force field, as it enables a set of parameters developed and tested on a relatively small number of cases to be applied to a much wider range of molecules.

The common force fields in use today were developed by using high-level quantum calculations and/or fitting experimental data. Only if constructed and parameterized correctly will the energy models generate reliable structural predictions. There are a number of complete sets of energy parameters available, which describe different interactions between, for example, all atoms in a protein or nucleotide. These force fields have their roots on the pioneering work of Momany et al. [52, 53] and Lifson/Hageler et al. [54]. Momany and his coworkers developed the Emperical Comformational Energy program for Peptides Potential (ECEPP) by starting from accurate geometric data for all amino acids while keeping bond lengths and bond angles fixed and terms optimized to crystal packing data, while, Lifson and Hagler developed an empirical data set for amides and carboxylic acids where a limited set of parameters (including partial charges, but without special hydrogen parameters) was determined by least squares fitting to crystal structures, lattice energies and a few dipole moments.

The force field used in this study is AMBER [55], an acronym for Assisted Model Building and Energy Refinement. It is a family of force fields for molecular dynamics of biomolecules originally developed by the Peter Kollman's group at the University of California, San Francisco and also refers to the molecular dynamics simulation package that implements these force fields.

The functional form of the AMBER force field [56] is:

$$V(r^N) = V_{bonded} + V_{nonb}$$

$$V_{bonded} = \sum_{bonds} K_b (d-d_o) + \sum_{angles} K_a (\theta - \theta_o) + 1/2 \sum_{dihedrals} K_d (1+\cos(m\Phi - \gamma)$$

$$V_{nonb} = \sum_{nonbo} (A/r^{12} - B/r^6 + q_i q_j / 4_{\Pi} \varepsilon_0 r_{ij})$$

V denotes the potential energy, which is a function of the positions (r) of N particles (usually atoms). The various contributions are schematically represented in Figure 7.

The first term in the equation, $\sum_{bonds} K_b (d-d_o)$, represents the energy between covalently bonded atoms. The two parameters that define the bond are the force constants $K_b$ and $d_o$. The interactions between a pair of bonded atoms is modeled by a harmonic potential that gives the increase in energy as the bond length d deviates from the reference, $d_o$. The force is a good approximation near the equilibrium bond length, but becomes increasingly poor as atoms separate.

The second term $\sum_{angles} K_a (\theta - \theta_o)$ is an energy summation over all valence angles in a molecule and is again modeled using a harmonic potential and requires two parameters; $K_a$ is a force constant and is an ideal bond angle while $\theta_o$ depends primarily on the hybridization of the central atom in the angle.

The third term $1/2 \sum_{dihedrals} K_d (1+\cos(m\Phi - \gamma)$ , represents the energy of twisting a bond due to bond order ( single or double bonds) and neighboring bond or lone pairs of electrons. It is a torsional potential and models how the energy changes as the functional

groups at either end of a bond rotate relative to each other. Each dihedral in the molecule requires three parameters to calculate the energy of the angle, a rotational barrier ($K_d$), number of local minima(m), and the value of the dihedral angle at which the energy is a minimum (related to $\gamma$).

The fourth and last contribution $\sum_{nonbo} (A/r^{12} - B/r^6 + q_i q_j/4\Pi\varepsilon_0 r_{ij})$ is the non-bonded term. It represents the non-bonded energy between all atom pairs that are in different molecules or that are in the same molecule but are separated by at least three bonds. The non bonded term can be decomposed into van der Waals interactions, modeled by using Lennard-Jones potential [57] and electrostatic energies, modeled using a Coulomb potential [58, 59] term where r is the inter-atomic distance between two ions, $q_i$ and $q_2$ the electric charges in coulombs carried by atoms 1 and 2 respectively, and $\varepsilon_0$ is the electrical permittivity of space. The electrostatic interaction is neglected for atoms sharing a common bond or bond angle. The parameter A accounts for intermolecular repulsion at short distances. The parameter B accounts for attractive dispersion forces.

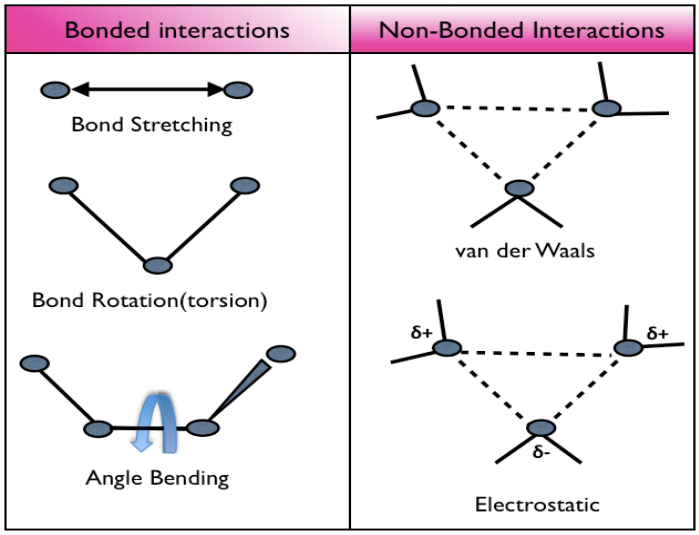


**Figure 7:** Schematic representation of the four key contributions to a molecular mechanics force field; bond stretching, angle bending, torsional terms and non-bonded interactions.

### 1.2.3 Molecular Dynamics Simulation

The MD approach is simple in principle. A system is stimulated in motion by following molecular configurations in time according to Newton's equation of motion (F=ma) under the influence of a specified force field (which as stated earlier is essentially the mechanical representation of the system and assumes simple, pair wise –additive potentials for the system, that expresses how the composite atoms stretch, vibrate, and rotate about the bonds in response to intramolecular and intermolecular forces). The result is a trajectory that specifies how the positions and velocities of the particles in the system vary with time.

A MD trajectory consists of three essential parts: initialization, equilibration, and production. An initial configuration of the system is first established. An equilibration phase is then performed, during which the system evolves from the initial configuration. Thermodynamic and structural properties are monitored during the equilibration until stability is achieved. At the end of the equilibration, the production phase commences. It is during the production phase that simple properties (e.g. energy, temperature, pressure and density) of the system are calculated. At regular intervals, the configuration of the system is outputted to a disk file. Finally the simulation is analyzed; properties not calculated during the simulation are determined and the configuration examined both to discover how the structure of the system changed and check for any unusual behavior that might indicate a problem with the simulation.

The first stage, initialization, involves specifying the initial coordinates and velocities for the solute macromolecule and for the solvent and ion atom. The initial

coordinates are either available from experiment (e.g. crystal structures) or can be acquired through homology modeling. Not surprisingly, the starting coordinates collected from experiment and homology modeling generally does not correspond to a minimum in potential energy. This is due to methods used to obtain crystal structures neglecting effect of solution to structure, thus not encouraging stable states of molecular system that correspond to local minimum energy. In addition, choice of force field used to represent crystal structure also affects net potential energy. Minimization (further refinement) is therefore usually done to relax strained contacts. First order minimization algorithm (energy minimization techniques) that are frequently used in molecular modeling to minimize the potential energy include steepest descent and conjugate gradient minimization techniques [60]. These gradually change the coordinates of the atoms as they move the system closer and closer to a minimum energy. The starting point for each iteration is the molecular configuration obtained from the previous step. For the first iteration, the starting point is the initial configuration of the system provided by the user. Energy minimizations generate individual minimum energy configurations of a system. In some cases, the information provided by energy minimization can be sufficient to predict accurately the properties of a system. However, this is possible only for relatively small molecules of small molecular assemblies in the gas phase where minimum configurations on energy can be identified and statistical mechanical formulae can be used to derive a partition function from which thermodynamic properties can be calculated. To study more complex systems (like the current HA-AB system), energy minimization techniques prepare the system for other types of calculations (like

molecular dynamics) by relieving any unfavorable interactions in the initial configuration of the system.

Many simulations of macromolecules were initially simulated *in vacuo* due to computer limitation, but with the realization that most macromolecules are naturally hydrated and hydration plays a big role of how they function, there was a need to come up with a way to get the correct balance between water and water solute interaction in the calculation. This problem was solved with the increase in the power of modern computers, which has allowed inclusion of explicit solvent (water molecules and counter ions) around the protein. Choosing the best water model has been a major interest of research in itself because of the problem of finding a model that can represent the anomalous properties of water in solid, liquid and gas phases. The current commonly used models are the SPC mode [61] and TIP3P and TIP4P model [62]. In TIP3P model (used in this study), three sites are used for the electrostatic interactions; the partial positive charges are partially balanced by an appropriate negative charge located on the oxygen atom. The van der Waals interaction between two water molecules is computed using a Lennard-Jones function [57] with just a single interaction point per molecule centered on the oxygen atom; no van der Waals interactions involving the hydrogen atoms are calculated.


**Figure 8:** TIP3P water model. It contains three simple interaction sites corresponding to the three atoms of the water molecule. Each atom gets assigned a point charge, and the oxygen atom also gets the Lennard-Jones parameters.

If crystallographic data is used, before continuing with the molecular dynamics, other factors need to be considered in the initial stages. For example, since the three-dimensional structures of biological macromolecules are stabilized by the presence of hydrogen bonds between secondary structural elements in proteins and between the bases in DNA and these are usually not included in crystallographic data, they need to be added in order to do a more accurate simulation. In addition, even though the coordinates of some of the solvent molecules may be known, it is necessary to add other solvent molecules to give the appropriate solvent density. This is usually done using the LEaP program in AMBER or other online programs, for example H++ [63-65].

Design of molecular dynamics simulations should also take into account available computational power. Simulation size (number of particles), time step and total time duration must be selected so that the calculation can be completed within a reasonable time period, but long enough to be relevant to the time scales of the natural process being studied. The spanning times related to protein dynamics most commonly indicated in literature vary between several nanoseconds to several microseconds, which take several CPU days or years. Parallel algorithms are sometimes used that allow the load to be distributed between CPU's. Computational costs can also be reduced by employing electrostatic methods such as the particle mesh Ewald [66], good spherical cut off techniques or using algorithms such as SHAKE [67], which fix the vibrations of the fastest atoms (e.g. hydrogen) into place. When a cut off is employed, the interactions between all pairs of atoms that are further apart than the cut off value are set to zero.

After doing the MD simulation, there are various different computational techniques available for use depending on properties to be studied to further predict the properties of a system like structure prediction, conformational analysis, sequence analysis, protein folding, solvation energies, and free energies.

### 1.2.4 Free Energy Calculations

The free energy is often considered the most important quantity in thermodynamics. The free energy is usually expressed as the Helmholtz function, A (used for a system with a constant number of particles, temperature and volume – constant NVT) [68] or the Gibbs function, G ( used for a system with constant number of particles, temperature and pressure - constant NPT ) [69]. Most experiments are conducted under conditions of constant temperature and pressure, where the Gibbs function is the appropriate free energy quantity.

Various techniques have been used to predict relative binding free energies of protein-protein association [70-73]. The Molecular Mechanics Poisson Boltzmann/ Generalized Born surface area (MM-PB(GB)SA) [74, 75] method was chosen for this study as other methods were not considered conducive as they suffer from long computational times in addition to potential technical difficulties associated with creation and annihilation of atoms. The MM–PB(GB)SA method is a recently introduced approach to predict free binding energies of a complex of molecules in solution. The method has been applied successfully to study protein-peptide, protein-protein, protein ligand, protein nucleic acid, nucleic acid-ligand interactions, processes such as protein folding, and the conformation- dependent free energies of nucleic acid. More specifically, the method has

been used to predict, the relative stabilities of A-DNA and B-DNA, effects of alanine mutations on protein-protein interactions [76], and binding of steroids to anti-progesterone and anti-testosterone antibodies [77].

MM-PB(GB)SA approximates free energies using a Poisson-Boltzmann (PB) continuum or generalized Born (GB) representation of the solvent together with a surface area-dependent term and molecular mechanics energies using snapshots from MD or Monte Carlo methods (not discussed) simulations generated with continuum solvent approaches. All snap shots collected are stored in AMBER format and the molecular energies are determined with the *anal* program from AMBER, representing internal energy (bond, angle and dihedral) and van der Waals electrostatic interactions.

Free energy calculations performed by using the MM-PB(GB)SA approach are described by Srinivasan et. al [78] and discussed in more detail in the method section. To use the MM-PB (GB)SA method, the user has to supply:

1) Trajectory files from MD
2) The input files for the anal program
3) The topology files for all interacting molecules without waters or their mutants, depending on the purpose of the run
4) The files with the charge and van der Waals parameters

## 1.3 Research Question/Hypothesis

As stated earlier, the inhibition of binding of the influenza HA to the sialic cell receptor is an effective way of preventing infection as it contains a sialic acid cell receptor-binding site. Mutations at these positions causes reduction in antibody affinity, allowing viruses to escape from neutralization without imposing selective pressure on residues of the receptor-binding site that could compromise its activity. As a complementary approach to experiments, molecular modeling and molecular dynamics simulations using the AMBER force field and MM-PB(GB)SA free energy calculation method are used to collect information regarding the trend and effect of single and multiple mutations of HA to antibody binding specificity of seven different isolates of the influenza virus over a range of 20 years. Virus strains from later years were expected to have similar or increased affinity to earlier virus-specific antibodies when they no longer posed a primary threat to the virus survival.

# METHODS

## 2.1 System set up for MD simulation

### 2.1.1 Sequence and Data Base Searches and Model Setup

The original influenza isolate chosen to be studied was based on the availability of a previously published paper[46] that described an X-ray crystallographic structure of an HA-AB complex. From a search of the Protein Data Bank (PDB) data base, initial atomic coordinates for the antibody complex were obtained from the X-ray crystal structure of a recombinant influenza strain containing A/Aichi/68 (H3N2) HA complexed with the Fab fragment of the anti-influenza neutralizing antibody (PDB ID 1EO8) by Fleury et al., determined to 2.8 Å resolution [46, 79]. No crystallographic waters were present in this crystal structure

The Influenza Virus Resource website [80] was used to search for sequences from the NCBI Influenza Database of other Hong Kong H3N2 virus isolates. A phylogenetic dendrogram was then built to enable the selection of influenza viruses with the least similarity and most differences between the sequences. Six influenza viruses were selected: A/Hong Kong/l/68 (H3N2), A/Hong Kong/l/69 (H3N2), A/Hong Kong/l/74 (H3N2), A/Hong Kong/l/75 (H3N2), A/Hong Kong/l/85 (H3N2), A/Hong Kong/l/92 (H3N2) and A/Hong Kong/l/99 (H3N2).

To construct the three-dimensional models of the six HAs, a multiple sequence alignment of the viruses versus the A/Aichi/68 influenza HA of known structure was done using NCBI -BLASTP 2.2.20+ [81] so as to find areas where mutations had taken place. BLAST (Basic Local Alignment Search Tool) is an online program that finds

regions of local similarity between sequences, by comparing nucleotide or protein sequences to sequence databases and calculating the statistical significance of matches. All the HAs and their complexes were then constructed using Swiss-PdbViewer (Swiss-Pdb Viewer, also known as deep view, is an application that provides a user friendly interface that allows analysis and modeling of several proteins at the same time [82]) and RasMol (molecular visualization software [83]), from the initial coordinates of the A/Aichi/68 (H3N2) HA–AB crystal structure. Amino acids were added, deleted or replaced with residues manually to create mutated proteins. It was discovered that the original PDB had some errors in residue numbering in the antibody (number 32 was missing*).* A program was created by A. Poe (NMU Computer Science Department) to correct this so that the numbering for all the atoms was consistent.

### 2.1.2 Structure Refinement

A preliminary pKa calculation was performed for all the starting structures of HA and their complexes using the WebServer H++ [63, 65], which is an automated system that computes pK values of ionizable groups in macromolecules and adds missing hydrogen atoms according to the specified pH of the environment. The PDB files generated by the H++ program were further edited after they were generated so that the LEaP program of AMBER could correctly read them. Some of the changes that had to be made manually in the PDB files included altering three letter abbreviations of Histidine, Lysine and Cysteine. Histidine can exist either as the protonated species or as a neutral species with a hydrogen at the delta or epsilon position. For this reason, the Histidine unit/residue name is either HIP, HID, or HIE (but not HIS). The AMBER force fields also differentiate

between the residue Cysteine (CYS) and similar residues that participates in disulfide bridges, Cystine (CYX), substituting CYX for CYS in the sequence was thus done to ensure correct disulfide bond creation. Aspartic acid--protonated ASH, Cystine, S--S crosslink CYX, Histidine, delta H HID, Histidine, epsilon H HIE, Histidine, protonated HIP. A general rule of thumb that was followed was to keep editing the input PDB file until LEaP stopped reporting errors for all the structures.

Sodium ions/Chloride counter ions were then added to the HA-AB complex by using a coulombic grid potential [59] via the LEaP module of AMBER to maintain the system at charge neutrality. The grid potential positions the ions at the most energetically favorable position near the protein. The system was then solvated using equilibrated TIP3P water in a rectangular periodic box as a building block [62]. The complex had at least a 10 Å buffer in every direction of the box to permit substantial fluctuations of the conformation during the course of the MD simulation.

## 2.2 Evaluation by Molecular Dynamics

To fix any errors introduced in the model-building process, the Sander module in the AMBER 7 program package was used to carry out minimization and molecular dynamics. The program relaxes the structure by iteratively moving the atoms down the energy gradient until a sufficiently low average energy gradient is obtained.

Non–bond long range interactions were cut off at 12.0 Å and three stages of minimizations performed to the water, hydrogens, amino acid side chains, and the entire

structure; so as to remove initial steric clashes. The whole system (including protein (HA alone or HA-AB complexes), water and hydrogen) was minimized with 200 steps of minimizations. The protein atoms were then fixed, to allow only the water to move and 500 steps of minimization done in a constant volume periodic boundary to relax the water. An additional 500 steps of minimization were then done on the whole system.

After minimization, the system was equilibrated in two stages. The protein molecules were first fixed and the water molecules of the solvated systems equilibrated using MD, and then the entire system equilibrated. Due to a problem with the AMBER program (the program kept shutting down), the first stage of equilibration was skipped on the HA-AB complex of 85, 92 and 99. At the beginning of each stage, the temperature of the system was increased gradually from 10 K to 300 K with a 5 ps time constant for heat bath coupling of the system. The target constant pressure of 1 atm was achieved and maintained by isotropic position scaling, a coupling algorithm used in AMBER. A 0.002 ps time step was used. The Berendsen temperature coupling algorithm [84] was used to maintain the system at its assigned temperature (300 K), with a scaling factor time constant. All covalent bonds in the protein and in water were represented by constraints that kept the bond distances to the proper chemicals values. The SHAKE algorithm [67] was used to fix the length of bonds containing a hydrogen atom to their equilibrium values. Constraints were applied to all bonds involving hydrogen atoms of 0.2 Å. The particle mesh Ewald [66] method was used to treat long range electrostatic interaction. To minimize computational expense, long-ranged non-bonded interactions (van der Waals interactions) were calculated out to a 12 Å residue-base cutoff distance.

Before proceeding with the production MD run, temperature, density and total energy versus time graphs were constructed to verify that the system had equilibrated. From the figures, it was clear that the energy, temperature and density had all clearly converged by the end of the equilibration period, and thus as a final step to the MD simulations, 600 ps production simulation were performed at constant pressure and temperature (300 K) after the system was equilibrated. A snapshot of the trajectory was stored every ps (500 time steps) for later analysis. To allow for sufficient equilibration, only the latter 500-ps of each 600-ps simulation were used for subsequent analysis of the dynamics.

In all calculations, the AMBER 94 force field [55, 56] was used. The MD simulation was performed using the AMBER suite of programs, version 7.0 [85]. Apart from what is mentioned above, all other default settings were used in the simulations. Illustrations of proteins were generated with the RasMol program [83]. The different steps of the simulations were performed partly on a parallel 2.8 GHz Pentium IV processor personal IBM laptop computer, partly on a 2.2 GHZ Intel core 2 Duo Mac OS X laptop and partly on a Beowulf cluster composed of a 32 node dual processor.

## 2.3 Energy Analysis: MM–PB (GB) SA Calculations

The MM-PB(GB)SA [75] method as implemented in AMBER 7 was used to calculate the binding energy for non-covalent association of the influenza viruses and the antibody. In the MM-PB(GB)SA method the average binding free energy ($\Delta G_{\text{bind}}$) is calculated as shown in Figure 9.

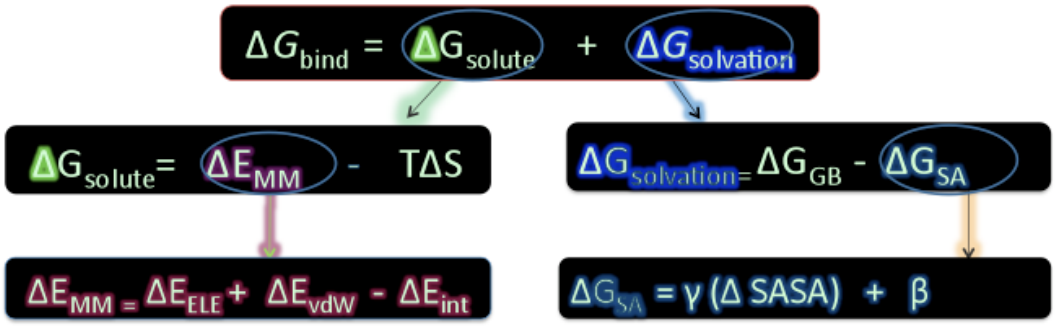


**Figure 9:** The binding free energy ($\Delta G_{bind}$) as defined by calculation from solute ($\Delta G_{solute}$) and solvent ($\Delta G_{solvation}$) contributions. Total molecular mechanical energy of the solute ($\Delta E_{MM}$) represents average interaction energies obtained from performing calculations on ensemble of uncorrelated snapshots from the equilibration MD; Electrostatic energy calculated from the Coulomb potential ($\Delta E_{ELE}$); van der Waals distance-dependent interaction energy calculated from the Lennard-Jones potential ($\Delta E_{VDW}$); Internal energy from bonds, angles and torsions ($\Delta E_{INT}$,); Electrostatic solvation contribution from solving the Generalized Born equation ($\Delta G_{GB}$) or Poisson Boltzmann equation ($\Delta G_{PB}$) of MM (GB/PBSA) in Amber 7; $\Delta G_{SA}$ is an empirical term for the non-electrostatic contribution. γ is a surface tension parameter, and was set to 0.00542 00542 kcal $mol^{-1}$ $A^{-2}$ for PB and 0072 kcal $mol^{-1}$ $A^{-2}$ for GB ; SASA is the solvent-accessible surface area determined by Sander's LCPO method in Amber and β is a parameterized value, set to 0.92 kcal $mol^{-1}$ for PB and 0 for GB.

As shown in Figure 9, the free energy of any binding process may be divided into a contribution from the solute and a contribution from the solvent:

$$G(X) = G_{solute}(X) + G_{solvation}(X)$$

The free energy contribution from the solvent and dissolved ions is calculated by solving the linearized PB or GB equation for each of the three states, unbound components and complex (gives the electrostatic contribution to the solvation free energy) and adding an empirical term for the non electrostatic contribution, calculated from the solvent–accessible surface area (SASA). This may be expressed as follows:

$$\Delta G_{solvation}(X) = \Delta G_{GB}(X) + \Delta G_{SA}(X)$$

$\Delta G_{GB}$ is the electrostatic contribution, obtained from PB or GB method.

$\Delta G_{SA}(X)$ is the non-electrostatic contribution calculated by

$$\Delta G_{SA}(X) = \gamma(\Delta SASA) + \beta$$

$\gamma$ is a surface tension parameter, and was set to 0.00542 kcal $mol^{-1}$ $Å^{-2}$ for PB and 0.0072 kcal $mol^{-1}$ $Å^{-2}$ for GB. *SASA* (*X*) is the solvent-accessible surface area of molecule *X*, determined by Sander's (Simulated annealing with NMR-derived energy restraints. This allows for NMR refinement based on NOE-derived distance restraints, torsion angle restraints, and penalty functions based on chemical shifts and NOESY volumes) Linear Combinations of Pairwise Overlaps (LCPO) method [86]. In this experiment $\beta$ is a parameterized value, set to 0.92 kcal $mol^{-1}$ for PB and 0 for GB.

Calculating the average interaction energy between the receptor and the ligand and taking the entropy change upon binding into account, if necessary, is how the free energy contribution from the solute is obtained. This may be expressed as follows

$$\Delta G_{solute} = \Delta E_{MM} - T\Delta S$$

$$\Delta E_{MM} = \Delta E_{ELE} + \Delta E_{VDW} + \Delta E_{INT}$$

$\Delta E_{MM}$ represents the average interaction energies obtained from performing calculations on an ensemble of uncorrelated snapshots collected from an equilibrated molecular dynamics (MD) simulation where $\Delta E_{ELE}$ is the electrostatic energy calculated from the Coulomb potential; $\Delta E_{VDW}$ is the van der Waals distance-dependent interaction energy calculated from the Lennard-Jones potential; $\Delta E_{INT}$ is the internal energy due to

bonds, angles and torsions. The entropy contribution (ΔS) can be found by performing normal mode analysis on the three species. However, entropy contributions are neglected in this case as normal mode analysis calculations are computationally expensive and tend to have a large margin of error that introduces significant uncertainty in the result. Neglecting entropy contributions has been found to still produce fairly accurate results when only a comparison of states of similar entropy is desired such as two ligands binding to the same protein [74].

In principle, the calculation of the binding free energy described above would require three independent MD simulations of the complex and both individual protein (HA and antibody). This protocol is referred to as "separate trajectories" or " three trajectories" (3T) by Gohlke and Case [74]. However an assumption was made that no significant conformational changes occur upon binding i.e. structural adaptation is negligible and the snap shots for all three species can be obtained from the single trajectory carried out on the complex by separating the complex into its constituent parts (Single Trajectory approach).

### 2.3.1 Generate Snapshots

To extract snapshots (without the water) from production runs for use in the MM-PBSA [75] calculation, the mm_pbsa.pl script (provided in AMBER 7), that automates this extraction process was used. Snapshots were generated for the complex as well as the receptor by itself and the ligand by itself from the MD trajectory file for subsequent MM-PB(GB)SA analysis. The input file used was the mm-pbsa.in (provided in AMBER 7). This input file specifies which atoms are part of the receptor, ligand and complex as well as specifying the total number of snapshots in the trajectories, the stride length, the names of the trajectory files and the prmtop files corresponding to the unsolvated structures. Prmtop files are topology files of the proteins that were created using LEaP after stripping water from solvated PDB files. After executing the mm_pbsa.pl command, approximately sixty equally spaced snapshots files were generated for the complex, receptor and ligand, from the trajectory file. As noted above, it is preferable that the snapshots are uncorrelated which is more likely if they are separated in time. The amount of snapshots (sixty) was selected so as to complete the calculation in an average of three days for each complex.

### 2.3.2 MM-PB(GB)SA Binding Energy Calculation

Starting with the snapshots extracted, the interaction energy and solvation free energy for complex, receptor and ligand were calculated and results averaged to obtain an estimate of the binding free energy. The binding energy calculation for the process A + B → AB i.e. the association of HA with AB to form the HA-AB complex was done using both the MM-PBSA method and the MM-GBSA method for comparison. This was accomplished by editing the input file for mm_pbsa.pl. Various portions of the input file specify which calculations to run, on which files to run them and any special parameters necessary to calculate the different contributions to the binding free energy. The ligand, receptor, and complex were all turned on (specified to be included in calculations), and the terms MM (Molecular Mechanics), GB, PB and MS(Molsurf) were computed for each species (see below).

MM - Calculation of gas-phase energies using Sander

GB - Calculation of solvation free energies using the GB models in Sander

PB - Calculation of solvation free energies using Delphi (The PB method called by the MM-PBSA, is implemented in the program Delphi). The Delphi program is not distributed with AMBER and therefore was installed locally [87-90].

MS - Calculation of nonpolar contributions to solvation using Molsurf. MolSurf is a program for the generation of molecular properties. MS = 0, in this study, so nonpolar contributions were instead calculated with the LCPO method in Sander.

When finished, as expected, five output files were produced: binding_energy.log, snapshot_statistics.out, snapshot_com.all.out, snapshot_rec.all.out, snapshot_lig.all.out.

The all.out files gave the individual energy contributions for each of the snapshots for each of the proteins (for the HA, antibody and complex, respectively) while the statistics.out file contains the energy components for complex, HA and antibody separately and the final averaged binding energy (under the category DELTA). A calculation of the entropy contribution to binding was not done, so strictly speaking the results of this study was not a true free energy value but could be used to compare similar systems. The log file just indicated the success of the completed calculation.

# RESULTS

Six mutated influenza HAs and their antibody complexes were successfully constructed and their change in binding energies calculated using the MM-PB(GB)SA method. The six, A/Hong Kong H3N2 strains chosen were from 1969, 1974, 1975, 1985, 1992, and 1999. They were selected to give the widest possible spread of variant types as shown by their dissimilarity to the original A/Hong Kong/l/68 (H3N2)-PDB (1EO8). This was based on a Hong Kong influenza phylogenetic tree (Figure 10) that demonstrated lack of similarity between the chosen influenza strains: A/Hong Kong/l/68 (H3N2), A/Hong Kong/l/69 (H3N2), A/Hong Kong/l/74 (H3N2), A/Hong Kong/l/75 (H3N2), A/Hong Kong/l/85 (H3N2), A/Hong Kong/l/92 (H3N2), A/Hong Kong/l/99 (H3N2). The 1999 HA was the most disimilar to the 1968 HA. Their HA-AB complexes were constructed using the 1968 HA X-ray structure as a template.

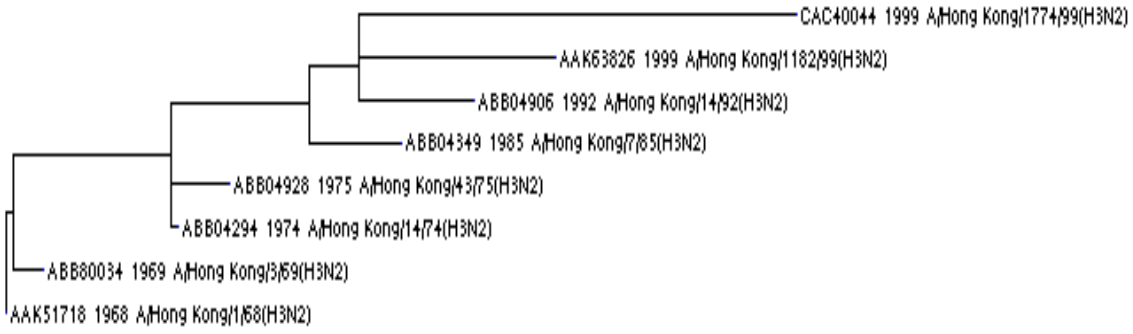


**Figure 10:** Phylogenetic relationship between the seven Hong Kong HAs depicting an evolutionary relationship and showing the lack of similarity between the chosen influenza isolates. The distance of one group from the other groups indicates the degree of relationship i.e., the most similar sequence pairs (Highest % sequence identity) are placed close together in a phylogenetic tree. Dissimilar pairs are placed furthest apart. The length of the horizontal lines is proportional to the number of amino acid differences. The 1999 HA is the most dissimilar to the 1968 HA.

To construct the three-dimensional models of the six HAs, a multiple sequence alignment of the HA protein sequences versus the A/Aichi/68 influenza virus of known structure was done so as to locate areas where mutations had taken place. The sequence alignment generated by NCBI–BLAST [81] is shown below in Figure 11 and the exact mutations in Table 1. As expected, high similarity is observed between all the proteins with the lowest sequence identity in a pair wise comparison found in antigenic sites of the influenza HA proteins [34, 91].

Most of the mutations were conservative in terms of polarity (e.g., 147, Asn → Ser, both polar amino acids or 137, Ile → Val, both hydrophobic). No conservation trend was seen in size of amino acid or acidity (e.g. some amino acids changed from larger to smaller one and others from basic to acidic or vice versa). The 1999 HA had the most mutations (41 amino acids changed from previous year) followed by 85 HA (23), 92 HA (14), 75 HA (10), 74 HA (9). The 1969 HA had the least (three).

**BLASTP 2.2.20+ RID: ZH0AE4G011N Query= gi|324132|gb|AAA43178.1| hemagglutinin precursor [Influenza A virus (A/Aichi/2/1968(H3N2))] Length=566**

**ALIGNMENTS**

```
Query     1    MKTIIALSYIFCLALGQDLPGNDNSTATLCLGHHAVPNGTLVKTITDDQIEVTNATELVQ  60
ABB80034  1    .............V.............................................  60
ABB04294  1    .............V..............................N.............  60
ABB04928  1    .............VFA............................N.............  60
ABB04349  1    .............VFA.K..........................N.............  60
ABB04906  1    ..........L..VFA.K..........................N.............  60
CAC40044  1    ............MV.......KG.N.......................V.........  60
Query     61   SSSTGKICNNPHRILDGIDCTLIDALLGDPHCDVFQNETWDLFVERSKAFSNCYPYDVPD  120
ABB80034  61   ................L.............................................  120
ABB04294  61   ..................N.............G....K.....................  120
ABB04928  61   ..................N.............G....K.....................  120
ABB04349  61   .....R..DS.......KN.............G....K.....................  120
ABB04906  61   .....R..DS.......KN.............G...KE.........Y.........  120
CAC40044  61   NL.M....S........AN.............G....K....I..............E  120
Query     121  YASLRSLVASSGTLEFITEGFTWTGVTQNGGSNACKRGPGSGFFSRLNWLTKSGSTYPVL  180
ABB80034  121  ..........................................................  180
ABB04294  121  ..................N...N.....................D.........Y.....  180
ABB04928  121  ..................N...N.I...................TD.........Y.....Q  180
ABB04349  121  ..................N...N.....S...Y.....SVNS........Y..E.K....  180
ABB04906  121  ..................N.D.N....A...D.Y.....SVKS........H..EYK..A.  180
CAC40044  121  H......I........VN.S.N.....................D.S.......Y...N...M.  180
Query     181  NVTMPNNDNFDKLYIWGIHHPSTNQEQTSLYVQASGRVTVSTRRSQQTIIPNIGSRPWVR  240
ABB80034  181  ......................V.....................................  240
ABB04294  181  ..................V.....D....N.............K.................  240
ABB04928  181  ..........S.......V.....DK...D.....K.....K.........V.......  240
ABB04349  181  .......GK........V.....DK...N...R...K.....K.....V...........  240
ABB04906  181  ........K........V.....DR.....IR...K.....K.....V...........  240
CAC40044  181  ......S.G........V.....DR..IN.....KI....K.........V........  240
Query     241  GLSSRISIYWTIVKPGDVLVINSNGNLIAPRGYFKMRTGKSSIMRSDAPIDTCISECITP  300
ABB80034  241  ..........................................T...........  300
ABB04294  241  ..................I.....................................G.........  300
ABB04928  241  ..................I.....................................G..S......  300
ABB04349  241  ..................I.L...T..........I.....................G..N......  300
ABB04906  241  ..................I.L...T..........I.....................G..S......  300
CAC40044  241  ..................I.I.S.....VH.....................E..S......  300
Query     301  NGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGMRNVPEKQTRGLFGAIAGFIENGWE  360
ABB80034  301  ............................................................  360
ABB04294  301  ...........................................I..............  360
ABB04928  301  ...........................................I..............  360
ABB04349  301  ........................R...................I..............  360
ABB04906  301  ..............R.......R...................I..............  360
CAC40044  301  ....................................I.......I..............  360
Query     361  GMIDGWYGFRHQNSEGTGQAADLKSTQAAIDQINGKLNRVIEKTNEKFHQIEKEFSEVEG  420
ABB80034  361  ............................................................  420
ABB04294  361  ............................................................  420
ABB04928  361  ............................................................  420
ABB04349  361  ..V.................................L.........................  420
ABB04906  361  ..V.................................L.........................  420
CAC40044  361  ..V.........................N..................................  420
Query     421  RIQDLEKYVEDTKIDLWSYNAELLVALENQHTIDLTDSEMNKLFEKTRRQLRENAEEMGN  480
ABB80034  421  ...................................................D....  480
ABB04294  421  ...................................................D....  480
ABB04928  421  ...................................................D....  480
ABB04349  421  ...................................................K.....D....  480
ABB04906  421  ...................................................K.....D....  480
CAC40044  421  ...................................................K.....D....  480
Query     481  GCFKIYHKCDNACIESIRNGTYDHDVYRDEALNNRFQIKGVELKSGYKDWILWISFAISC  540
ABB80034  481  ............................................................  540
ABB04294  481  ..............G.............................................  540
ABB04928  481  ..............G.............................................  540
ABB04349  481  ..............G.............................................  540
ABB04906  481  ..............G.............................................  540
CAC40044  481  ..L........S..D.........NE.............S....T..............  540
Query     541  FLLCVVLLGFIMWACQRGNIRCNICI  566
ABB80034  541  ..........................  566
ABB04294  541  ..............K.........  566
ABB04928  541  ..............K.........  566
ABB04349  541  ..............K.........  566
ABB04906  541  ..............K.........  566
CAC40044  541  ......W.........K........  565
```

**Figure 11**: The sequence alignment results of the six influenza HA viruses; 1969 HA (ABB80034), 1974 HA (ABB04294), 1985 HA (ABB04928), 1992 HA (ABB04349), and 1999 HA (ABB04906, CAC40044) with A/Aichi/68 (AAA43178) as generated by BLASTP 2.2.20+. Pink highlighted residues are the antigenic sites described by Wiley et. al [1].

**Table 1**: Color-coded table showing types of mutations. Yellow represents first time an amino acid mutates; green, amino acid mutated/reverts back to the original amino acid; orange, second mutation; red, third mutation from original; blue, amino acid mutates back to second mutation amino acid. The red number in brackets indicates number of amino acid changes in that year, when compared to previous year on table.

**COLOR CODED TABLE SHOWING TYPE AND NUMBER OF AMINO ACID MUTATIONS IN EACH HAEMAGLUTININ**

| | 68 | 69 | 74 | 75 | 85 | 92 | 99 | | | 68 | 69 | 74 | 75 | 85 | 92 | 99 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | (3) | (9) | (10) | (23) | (14) | (41) | | | | | | | | | |
| **13** | Leu | Leu | Leu | Leu | Leu | Leu | Met | | **162** | Gly | Gly | Gly | Gly | Ser | Ser | Ser |
| **14** | Ala | Val | Val | Val | Val | Val | Val | | **171** | Thr | Thr | Tyr | Tyr | Tyr | His | Tyr |
| **15** | Leu | Leu | Leu | Phe | Phe | Phe | Leu | | **175** | Ser | Ser | Ser | Ser | Ser | Tyr | Asn |
| **16** | Gly | Gly | Gly | Ala | Ala | Ala | Gly | | **176** | Thr | Thr | Thr | Thr | Lys | Lys | Thr |
| **18** | Asp | Asp | Asp | Asp | Lys | Lys | Asp | | **179** | Val | Val | Val | Val | Val | Ala | Met |
| **22** | Asn | Asn | Asn | Asn | Asn | Asn | Lys | | **180** | Leu | Leu | Leu | Gln | Leu | Leu | Leu |
| **23** | Asp | Asp | Asp | Asp | Asp | Asp | Gly | | **187** | Asn | Asn | Asn | Asn | Asn | Asn | Ser |
| **24** | Asn | Asn | Asn | Asn | Asn | Asn | Asn | | **188** | Asp | Asp | Asp | Asp | Gly | Asp | Asp |
| **25** | Ser | Ser | Ser | Ser | Ser | Ser | Asn | | **189** | Asn | Asn | Asn | Asn | Lys | Lys | Gly |
| **70** | Asn | Asn | Asn | Asn | Ser | Ser | Asn | | **190** | Phe | Phe | Phe | Ser | Phe | Phe | Phe |
| **78** | Ile | Leu | Ile | Ile | Lys | Lys | Ala | | **209** | Ser | Ser | Asn | Asp | Asn | Ser | Asn |
| **94** | Val | Val | Gly | Gly | Gly | Gly | Gly | | **213** | Gln | Gln | Gln | Gln | Arg | Arg | Gln |
| **98** | Glu | Glu | Glu | Glu | Glu | Lys | Glu | | **217** | Arg | Arg | Arg | Lys | Lys | Lys | Lys |
| **99** | Thr | Thr | Lys | Lys | Lys | Glu | Lys | | **218** | Val | Val | Val | Val | Val | Val | Ile |
| **104** | Val | Val | Val | Val | Val | Val | Ile | | **223** | Arg | Arg | Lys | Lys | Lys | Lys | Lys |
| **110** | Phe | Phe | Phe | Phe | Phe | Tyr | Phe | | **229** | Ile | Ile | Ile | Ile | Val | Val | Ile |
| **121** | Tyr | Tyr | Tyr | Tyr | Tyr | Tyr | His | | **264** | Asn | Asn | Asn | Asn | Thr | Thr | Asn |
| **137** | Ile | Ile | Ile | Ile | Ile | Ile | Val | | **276** | Met | Met | Met | Met | Ile | Ile | Val |
| **138** | Thr | Thr | Asn | Asn | Asn | Asn | Asn | | **277** | Arg | Arg | Arg | Arg | Arg | Arg | His |
| **140** | Gly | Gly | Gly | Gly | Gly | Asp | Ser | | **289** | Pro | Thr | Pro | Pro | Pro | Pro | Pro |
| **142** | Thr | Thr | Asn | Asn | Asn | Asn | Asn | | **291** | Asp | Asp | Gly | Gly | Gly | Gly | Glu |
| **144** | Thr | Thr | Thr | Ile | Thr | Thr | Thr | | **294** | Ile | Ile | Ile | Ser | Asn | Ser | Ser |
| **147** | Thr | Thr | Thr | Thr | Thr | Ala | Thr | | **492** | Ala | Ala | Ala | Ala | Ala | Ala | Ser |
| **148** | Gln | Gln | Gln | Gln | Gln | Gln | Gln | | **520** | Gly | Gly | Gly | Gly | Gly | Gly | Ser |
| **149** | Asn | Asn | Asn | Asn | Ser | Asn | Asn | | **315** | Lys | Lys | Lys | Lys | Lys | Arg | Lys |
| **153** | Asn | Asn | Asn | Asn | Tyr | Tyr | Asn | | **323** | Lys | Lys | Lys | Lys | Arg | Arg | Lys |
| **159** | Pro | Pro | Pro | Thr | Ser | Ser | Pro | | **324** | Tyr | Tyr | Tyr | Tyr | Tyr | Tyr | Tyr |
| **160** | Gly | Gly | Asp | Asp | Val | Val | Asp | | **363** | Ile | Ile | Ile | Ile | Val | Val | Val |
| **161** | Ser | Ser | Ser | Ser | Asn | Lys | Ser | | **547** | Leu | Leu | Leu | Leu | Leu | Leu | Trp |

All the HAs and their complexes were constructed using Deep view and RasMol from the initial coordinates of the A/Hong Kong/l/68 (H3N2)-PDB (1EO8) crystal structure. As can be seen in Figure 12, the vast majority of mutations occur in the distal sialic binding domain rather than in the long helical region. It makes sense that the mutations would cluster in the antigenic region since changes to the protein structure in these regions are likely to allow for escape from antibody. The bottom box in Figure 12 is a representation of the simulation box of the full system of the 1968 HA-AB complex. The protein had at least a 10 Å buffer in every direction of the rectangular periodic box.

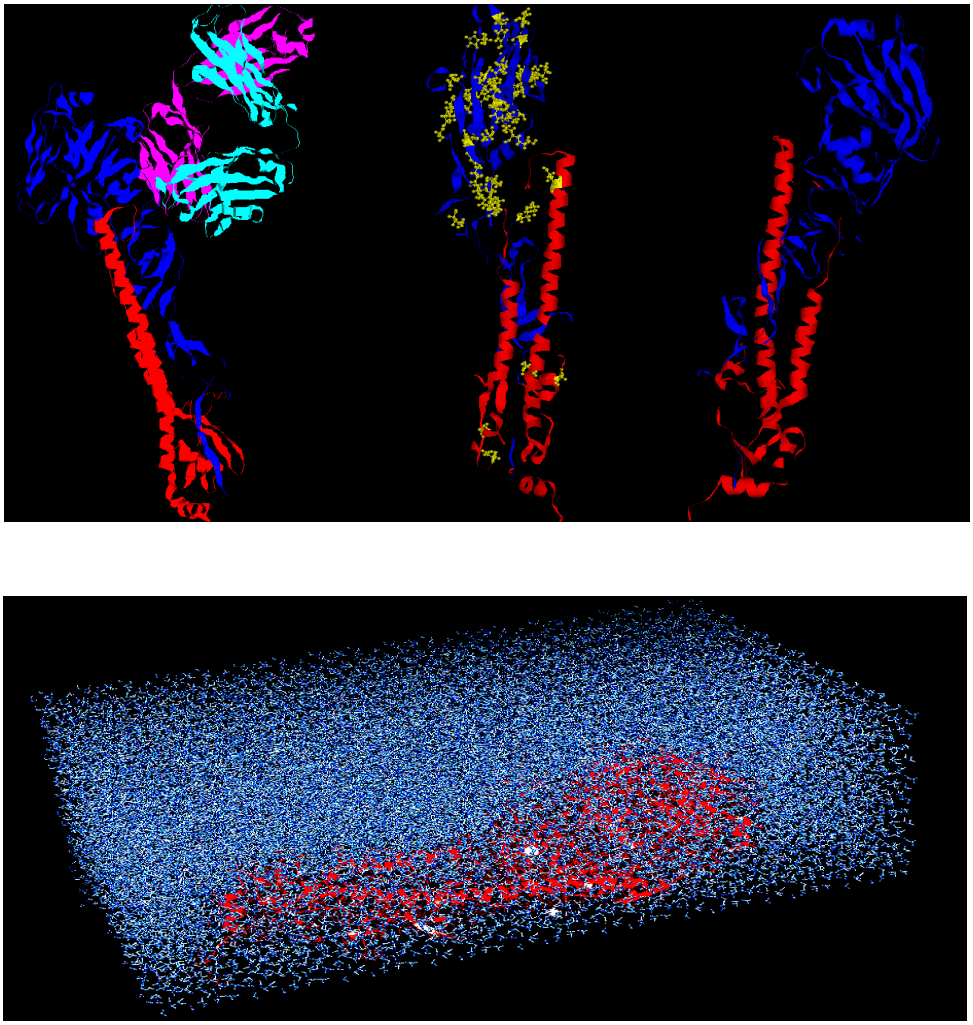

**Figure 12**: Ribbon representations of mono-subunits of the homo-trimmeric, hemagglutinin, glycoprotein of the original 1968 HA, mutated 1999 HA, the 1968 HA-AB complex (from right to left). The blue and red ribbons show the two distinct chains found in a mono-subunit of HA. The magenta and cyan ribbons in HA-AB complex represent the light and heavy chain, respectively. The yellow amino acids in the 1999 HA represent the mutated residues. The bottom box is a representation of the simulation box of the full system of the 1968 HA-AB complex (the complex is represented by red ribbons and water by blue wires). The total system size is 188605 atoms.

The preliminary pI calculations performed for both starting structures using the WebServer H++, Sodium ions/Chloride counter ions added to the HA/HA-AB complex and the total amount of water molecules added are shown in Table 2 below. The antibody seems to be positively charged as it causes the complex to be more positive when

compared to the HA alone. E.g., for the 1968 HA, charge at pH 6.5 is -3, when antibody is added to form complex, total charge is +3. Earlier HAs have a more negative total charge at pH 6.5 and lower isolectric point. The isolectric point and the total charge at pH 6.5 increases with year progression (lowest 1968, highest 1999). Interestingly, in the 1968 HA-AB complex, the HA is an anion while the AB is a cation, the proteins are thus complementary and encourage binding. However by 1985, both proteins are cations and binding would thus not be favorable. There also appears to be a change in protonation state of some of the antibody's amino acids upon HA, AB binding (HIS 196 and HIS 204); when in isolation the AB appears to have a + 8 charge, however, when in complex, it has a +6 charge.

**Table 2:** The preliminary pI of starting structures for both the HA and their complexes calculated using the WebServer H++ and total amount of waters and sodium/chloride ions added. Sodium ions/Chloride counter ions were added to the HA-AB complex using the columbic grid of the LEaP module of AMBER, and water added using an equilibrated TIP3P water in a rectangular periodic box as a building block. The antibody seems to be positively charged as it causes the complex to be positive when compared to the HA alone. Earlier HAs had a more negative total charge at pH 6.5 and lower isolectric points.

| **YR OF Hong Kong/ H3N2** | **HAEMAGLUTININ ALONE** | | | | **HAEMAGLUTIN-AB COMPLEX** | | | |
|---|---|---|---|---|---|---|---|---|
| | *pI* | ***Total charge at pH 6.5*** | ***N.O/Type of ions added in leap*** | ***Total number of $H_2O$ added in leap*** | *pI* | ***Total charge at pH 6.5*** | ***N.O/Type of ions added in leap*** | ***Total number of $H_2O$ added in leap*** |
| **68** | 6 | -3 | 3 Na+ | 24996 | 8 | 3 | 3 Cl- | 58169 |
| **69** | 6 | -3 | 3 Na+ | 24995 | 8 | 2 | 2 Cl- | 58168 |
| **74** | 7 | 0 | 0 | 25030 | 8 | 4 | 4 Cl- | 58091 |
| **75** | 7 | 1 | 1 Cl- | 25025 | 8 | 5 | 5 Cl- | 58086 |
| **85** | 9 | 6 | 6 Cl- | 26468 | 9 | 9 | 9 Cl- | 60018 |
| **92** | 8 | 5 | 5 Cl- | 26461 | 9 | 9 | 9 Cl- | 60001 |
| **99** | 9 | 6 | 6 Cl- | 26945 | 9 | 13 | 13 Cl- | 59978 |
| **AB** | 9 | 6 | 8 Cl- | 17255 | | | | |

To obtain reliable estimates of absolute binding free energies, the average values calculated by MM-PB(GB)SA had to be converged. After the equilibration run, density, temperature, pressure and total energy data for all HA complexes were extracted from the output files and combined to form continuous plots. It was confirmed that the systems were equilibrated as all the values appeared to converge and the temperature had reached the desired 300 K after 20 ps of equilibration (Figure 13). The average density was close to the expected 1.0 gm/ml, which is appropriate for a protein solution, and the average pressure was approximately 1 atm. Based on these plots it appears that the production runs were stable and well behaved.

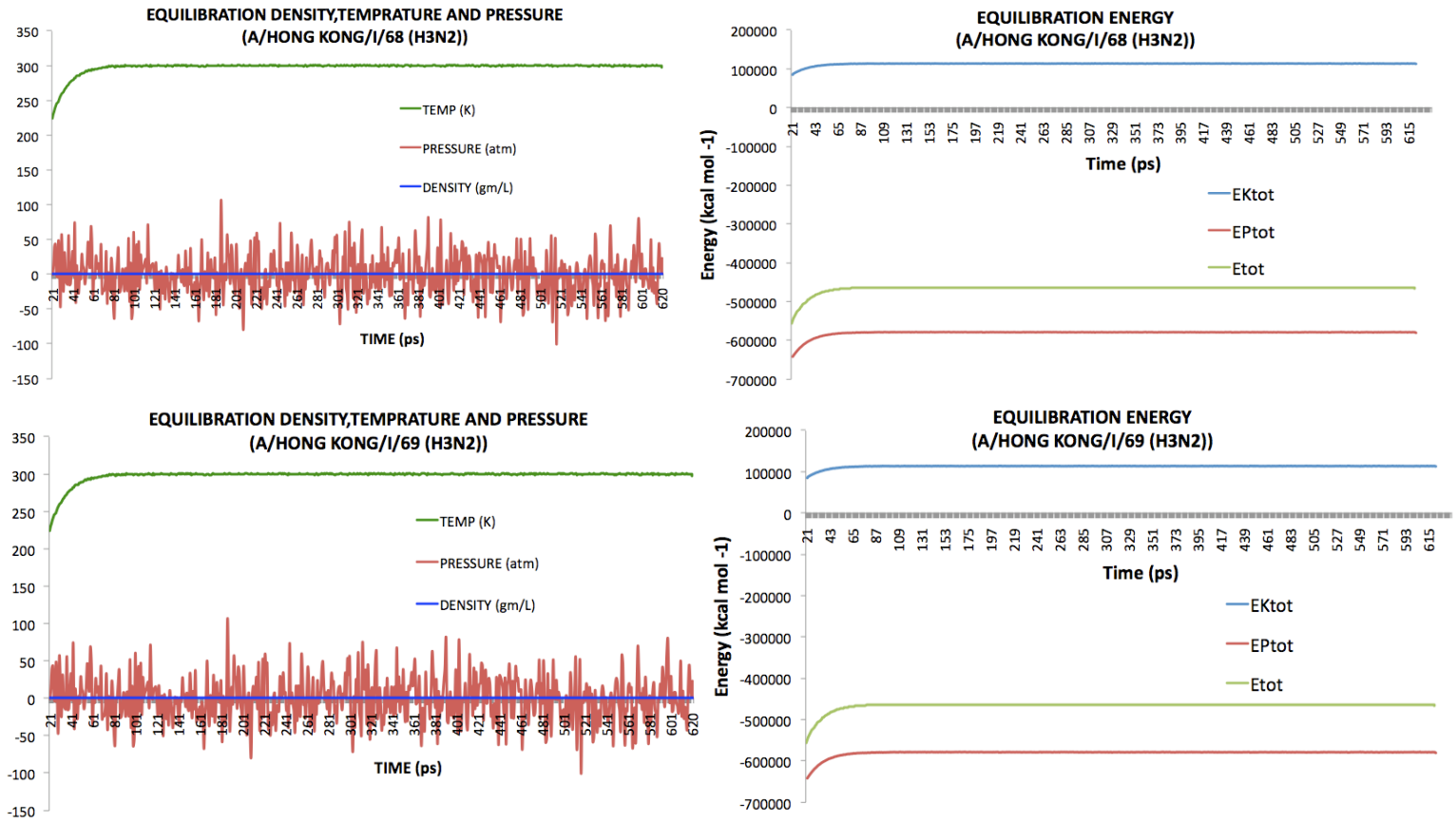


**Figure 13:** The plots of the density, pressure and temperature (left) and potential (Eptot), kinetic (Ektot) and total Etot energy (right), versus time, of the 1968 and 1969 HA-AB complex equilibration runs. The system was equilibrated in two stages (with the protein molecules fixed first and then the entire system). Temperature was increased gradually from 10K to 300 K while maintaining constant pressure. Berendsen algorithm was used to maintain system at its assigned temperature (300K). The graphs show that the systems had been brought safely from 10K up to 300 K after 20 ps of run time and the density of the water box adjusted to ~1.0 grams/L. The values on the Y-axis on the graphs on the right represent the temprature (K), pressure (atm) and density (gm/L), depending on what property is being followed. The values on the Y-axis on the graphs on the left represent energy (kcal/mol).

To further investigate the extent of equilibration, the root mean square deviation (RMSD) in C -α position compared to the starting structure was determined as a function of time. A typical RMSD plot is shown in Figure 14 for the simulation of the 1985 complex. It seems clear that there is a significant conformation shift occurring at about 250 ps. Based on the separate superposition of either just the antibody or just the HA (also shown in Figure 14), it is clear that the conformational shift is due to HA not the antibody. The B factor plot and RMSD plots of the 85 HA-AB complex seemed to indicate that the AB was more flexible than HA and while the AB had reached an equilibrated conformation state at around 100ps, HA had not (Figure 14). To further clarify the nature of the conformational shift, a B-factor plot was prepared and is included in Figure 14. For the HA, which represents residues 1- 566 in this plot, the most flexible region is around residue 322 which suggests a change in conformation of the long helical stalk which is far from the antigen binding site.

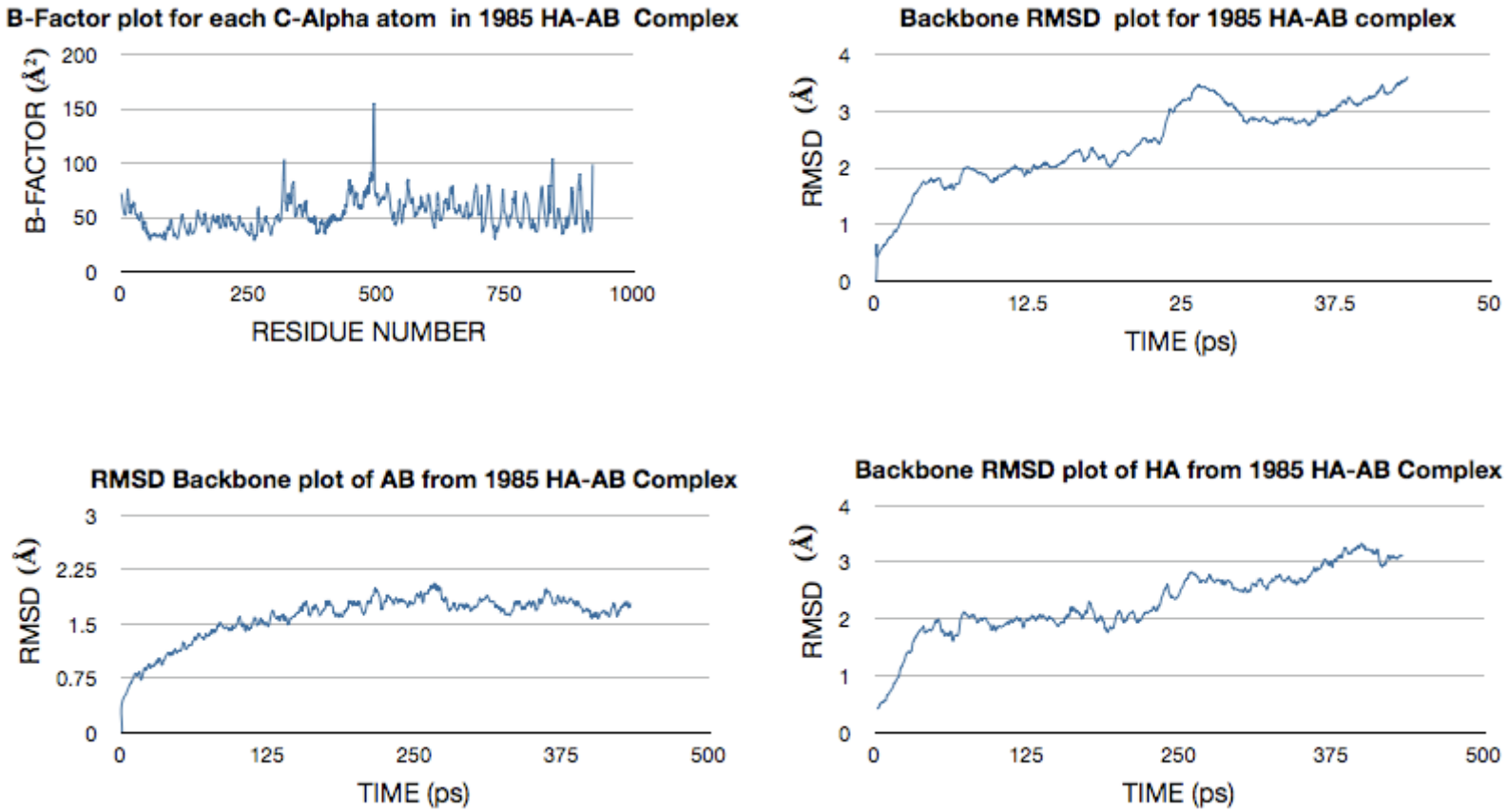


**Figure 14:** B-factor plot and backbone RMSD plots of the 1985 HA-AB complex. B-factor plot (top left) and back bone RMSD plots of the 1985 HA-AB complex (top right), AB from complex (bottom left) and HA from complex (bottom right). The B factor shows that the antibody is more flexible than the HA. The RMSD plots show that while the AB had definitely reached equilibrium, HA had not.

The MM-PB(GB)SA method was used to average contributions of gas-phase energies and solvation free energies calculated for snapshots of the HA–AB complex as well as the unbound components, which were extracted from MD trajectories. The binding free energy was obtained as the difference between the free energy of the complex and the individual molecules. In our case, all water molecules were stripped, as were all counter ions. Table 3 contains a free energy analysis showing energetic contributions to the binding free energy of HA-AB, obtained for 60 snapshots of equal spacing from the 500 snapshot MD trajectories of the HAs. Only 60 snapshots were selected to ensure the binding energy calculations finished in a reasonable time frame i.e. three days for each complex. The gas-phase energies included coulomb, van der Waals, and internal energies. The solvation free energy is obtained as sum of a solvent-accessible surface-dependent nonpolar contribution and a polar contribution from solving either the PB or GB equations (see, Figure 9). In all cases $\Delta E_{INT}$ is negligible which indicates conformational changes upon binding did not lead to any internal strain.

**Table 3:** Specific energy contributions to the binding free energy in kcal $mol^{-1}$. Numbers in brackets represent standard deviations. $\Delta E_{ELE}$, electrostatic molecular mechanical energy contribution calculated from the Coulomb potential; $\Delta E_{VDW}$, van der Waals distance-dependent interaction energy calculated from the Lennard-Jones potential; $\Delta E_{INT}$ is the internal energy from bonds, angles and torsions; $\Delta E_{MM}$, total molecular mechanical Energy (from generated snapshots); $\Delta G_{Solv}$, total solvation energy; $\Delta G_{GB}^{HCT}$, electrostatic solvation contribution from solving the Generalized Born equation, using the $GB^{HCT}$ model ( HCT refers to the GB method used (IGB=1). Further descriptions of this method may be found in the AMBER 7 manual); $\Delta G_{PB}$, electrostatic solvation contribution from solving the Poisson Boltzmann equation; $\Delta G_{Bind(GB)}/\Delta G_{Bind(PB}$, binding free energy from GB or PB method respectively.

| **Free Energy Analysis (Kcal $mol^{-1}$) Binding of Influenza HA to the 68 HA antibody using the MM-PB(GB)SA Method** | | | | | | | |
|---|---|---|---|---|---|---|---|
| | A/Hong Kong/l/68(H3) | A/Hong Kong/l/69(H3) | A/Hong Kong/l/74(H3) | A/Hong Kong/l/75(H3) | A/Hong Kong/l/85(H3) | A/Hong Kong/l/92(H3) | A/Hong Kong/l/99(H3) |
| $\Delta E_{ELE}$ | -272.11 31.06) | -271.4(22.97) | -19.06(21.63) | 22.56(30.95) | 98.07(27.96) | 80.54(17.78) | 297(12.61) |
| $\Delta E_{VDW}$ | -84.58 (3.97) | -83.32(6.21) | -80.68(4.84) | -87.98(5.27) | -94.61(5.53) | -85.77(2.41) | -95.14(5.62) |
| $\Delta E_{INT}$ | 0.01 (0.02) | 0.02(0.02) | 0.01(0.02) | 0.02(0.02) | 0.01(0.01) | 0.04(0.04) | 0 |
| $\Delta E_{MM}$ | -356.67(30.46) | -354.7(22.35) | -99.73(22.33) | -65.4(30.34) | -3.47(26.77) | -5.19(16.58) | -201.86(10.73) |
| | | | | | | | |
| $\Delta\Delta G_{np}$ | -10.00 (0.27) | -10.12(0.43) | -9.19(0.34) | -9.79(0.39) | -10.69(0.34) | -10.08(0.27) | -10.41(0.56) |
| $\Delta\Delta G_{PB}$ | 322.50(28.80) | 306(18.76) | 144.29(18.83) | 120.89(28.09) | 85.71(20.95) | 56.23(21.84) | -72.49(17.73) |
| $\Delta\Delta G_{Solv}$ | 312.51(28.76) | 295.87(18.63) | 135.1(18.68) | 111.09(27.86) | 75.02(20.79) | 46.15(21.63) | -82.9(17.56) |
| $\Delta\Delta G_{PB,Elec}$ | 50.39 (12.83) | 34.59(15.9) | 125.23(11.83) | 143.45(10.90 | 183.78(15.28) | 136.78(6.49) | 224.51(14.98) |
| $\Delta G_{Bind(PB)}$ | -44.17(11.75) | -58.83(13.04) | 35.37(11.35) | 45.69(10.53) | 78.49(14.24) | 40.96(8.72) | 118.96(10.7) |
| | | | | | | | |
| $\Delta\Delta G_{np,GB}^{HCT}$ | -12.06 (0.36) | -12.23(0.57) | -10.99(0.45) | -11.79(0.52) | -12.97(0.45) | -12.17(0.36) | -12.61(0.74) |
| $\Delta\Delta G_{GB}^{HCT}$ | 302.00(28.17) | 300.65(20.48) | 66.53(18.29) | 26.53(26.49) | -42.72(24.88) | -27.72(15.56) | -224.51(7.26) |
| $\Delta\Delta G_{Solv, GB}^{HCT}$ | 289.94(28.18) | 288.42(20.5) | 55.54(18.13) | 14.75(26.18) | -55.69(24.68) | -39.89(15.25) | -237.12(7.46) |
| $\Delta\Delta G_{GB,Elec}$ | 29.89 (5.27) | 29.25(7.87) | 47.47(5.98) | 49.1(6.28) | 55.35(6.14) | 52.82(2.34) | 72.49(7.28) |
| $\Delta G_{Bind(GB)}$ | -66.73 (5.49) | -66.28(5.3) | -44.19(6,54) | -50.65(5.64) | -52.22(4.89) | -45.08(3.08) | -35.25(5.44) |

To test the internal consistency of the computations, binding free energy calculations were repeated using only 5 snap shots of trajectories from 1969 HA. A comparison of the results with values obtained for 60 snapshots indicates essentially identical binding energy values ($64.88 \pm 6.1$ kcal $mol^{-1}$ vs. $66.28 \pm 5.3$ kcal $mol^{-1}$)

Figure 15 shows the different energy contributions to the binding energy. $\Delta E_{ELE}$ (electrostatic component from solute) is initially highly favorable for the 1968 HA complex but becomes very unfavorable in later years. This reflects the change in net charge of the HA as a result of the mutations. $\Delta E_{VDW}$ (van der Waals component) favors binding in all complexes and is fairly constant. $\Delta E_{INT}$ (internal component) and $\Delta G_{np}$ (non-polar/non electrostatic contributions to solvation energy) have very little contribution; $\Delta G_{GB}$ or $\Delta G_{PB}{}^{HCT}$ (electrostatic contribution to the solvation energy determined by the GB/PB equation), which are initially highly unfavorable, decrease with time thus partially balancing the unfavorable change in $\Delta E_{ELE}$. The 1992 HA (which also did not follow the overall increase in number of mutations trend when compared to the amount of mutations in the previous year) also showed a difference in trend in $\Delta E_{ELE}$ and $G_{GB}{}^{HCT}$. The value becomes slightly less positive for $\Delta E_{ELE}$ than the previous year while all the other HA values were more positive than the previous year. In addition, the value became slightly less negative for $G_{GB}{}^{HCT}$ than the 1985 HA while all the other $G_{GB}{}^{HCT}$ HA values were more positive than their prior year.

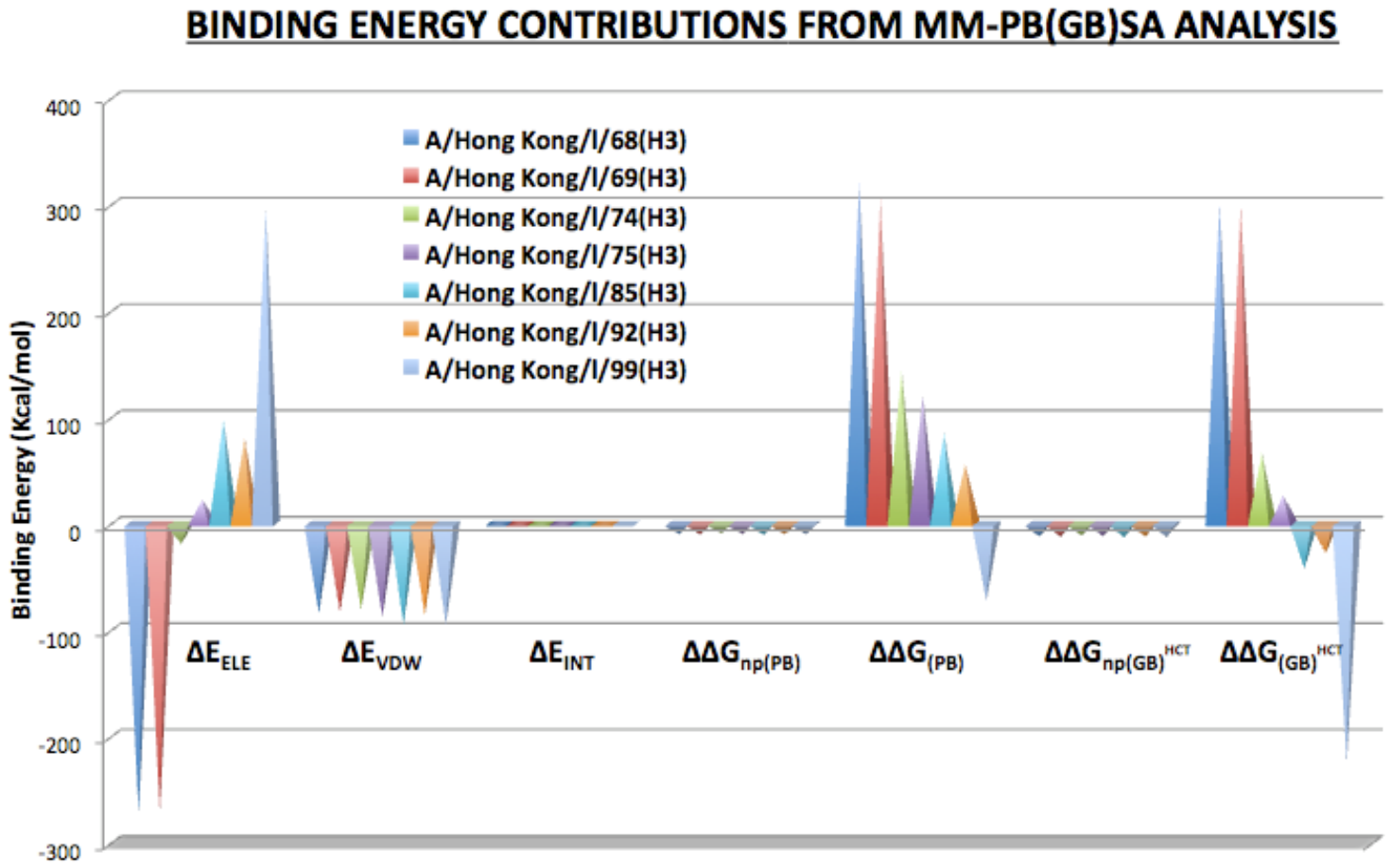


**Figure 15**: The different energy contributions to the binding energy. $\Delta E_{ELE}$ (electrostatic component) progressively decreased favor of binding; $\Delta E_{VDW}$ (van der Waals component) favors binding in all viruses; $\Delta E_{INT}$ (internal component) and $\Delta G_{np}$ (non-electrostatic contributions to solvation energy), very little contribution; $\Delta G_{GB}$/ $\Delta G_{PB}$ (electrostatic contribution to the solvation energy determined by the GB/PB equation), decreased contribution with time. HCT refers to the GB method used (IGB=1).

To show the trend of binding energy with year progression, as calculated by the MM-GBSA method, a graph of binding energy versus time (year) was constructed (Figure 16). The least negative binding energy (lowest antibody affinity) was displayed by the 1999 HA while the original HA from the 1968 HA had the highest negative change in binding energy (highest antibody affinity). This trend is consistent with the fact that the antibody is specific for the 1968 HA and that later years can presumably escape the antibody.

A graph was also constructed to show the trend of binding energy with year progression as calculated using the MM-PBSA method (Figure 17). A similar pattern of general decrease in binding affinity is seen with the least negative binding energy change (lowest antibody affinity) by the 1999 HA but the the most negative change in binding energy (highest antibody affinity) was by the 1969 HA mutant rather than the 1968 HA as might be expected. However, the difference in binding energy between 1968 and 1969 is barely larger than the standard deviation in the binding energy calculation and therefore the difference may not be significant.

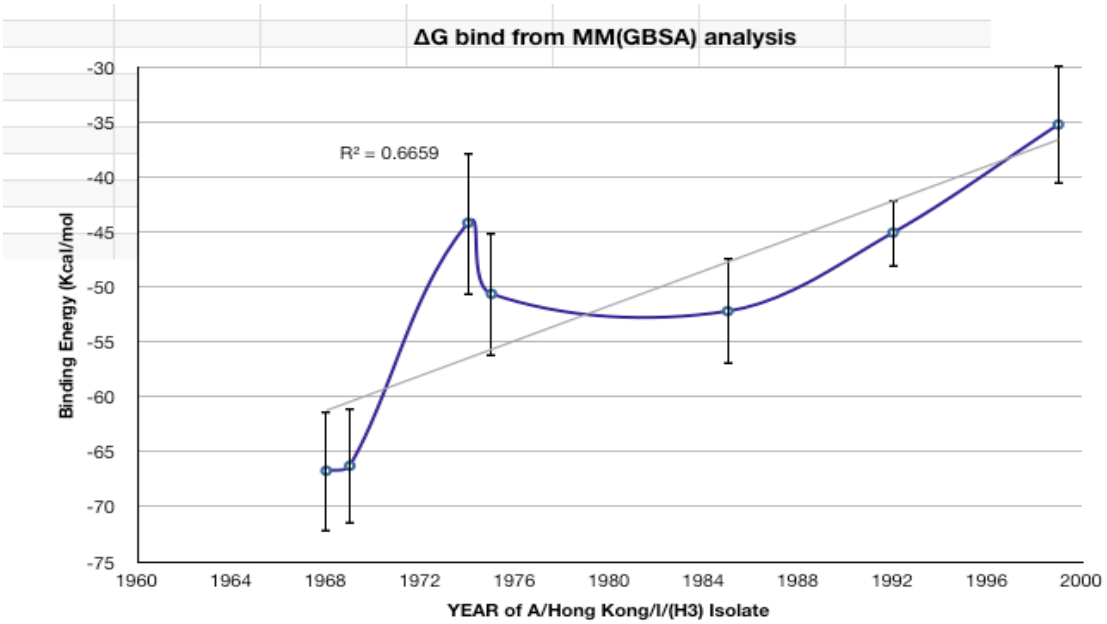


**Figure 16**: Graph showing change in free binding energy with increase in time/mutations, as calculated by MM-GBSA. The 1968HA had the highest negative free energy change (-66.73 ± 5.49 kcal $mol^{-1}$) while 1999 HA had the lowest (35.25 ± 5.44 kcal $mol^{-1}$). The error bars indicate standard deviation.

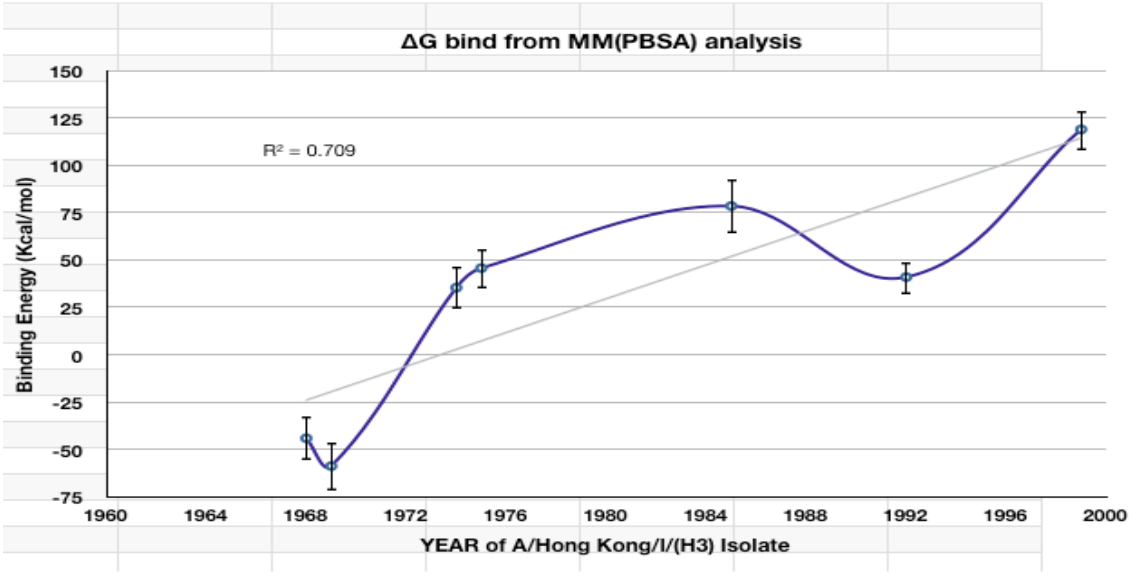


**Figure 17**: Graph showing the trend of binding energy with year progression as calculated using the MM – PBSA. The least negative binding energy change (lowest antibody affinity) was by the 1999 HA (118.96 ± 10.7 kcal $mol^{-1}$) while the 1969 HA (58.83 ± 13.83 kcal $mol^{-1}$) had the most negative change in binding energy (highest antibody affinity). The error bars indicate standard deviation.

# DISCUSSION

The sites of amino acid substitution in natural variants of HA from Hong Kong H3 subtype viruses that were analyzed were scattered throughout the HA. However, as suggested in various studies, most substitutions coincided with HA epitopes on proposed antigenic sites that surrounded the conserved host-receptor binding site, while mutations in the stalk were much less frequent [34]. Taking into account that the antibody footprint is larger than the receptor binding site, this is a very efficient way of preventing neutralization of the influenza virus as it reduces binding of antibodies that target this binding site without compromising its receptor binding function; which allows new virus subtypes to spread within non immune species. That most of the "fixed changes" (i.e. changes that were retained in HAs of isolated viruses in subsequent years), involved residues on the surface membrane–distal globular domain of HA that surround the receptor binding site whereas most of those that are not retained were found to be buried, was also the general trend found in the HAs of the seven strains of the influenza viruses analyzed. Fixed substitutions are thought to be naturally selected because they prevent antibody binding [35].

The fact that simulations of all the seven HA-AB systems were carried out with exactly the same computational procedures, allowed analytical comparisons to be made between the binding energetics of Hong Kong HA isolates to the 1968 HA antibody, over a period of twenty years. In line with the results of other studies [3, 44, 45, 92], binding free energy became less favorable with increase in mutations, from 1968 to 1999 (-66.73 to -35.25 kcal $mol^{-1}$). Fleury et al [46] also found that a HA wild type specific antibody still

bound to three escape mutant HAs, but with a much reduced affinity i.e. between $K_d$ $4 \times 10^{-6}$ M and $5 \times 10^{-7}$ M. This was significantly lower than its affinity for the wild-type HA ($K_d$ $10^{-9}$ M) [3]. However, the results of the MM-PBSA study show that there was no increase in antibody binding affinity after twenty years of mutations. The 1968 HA had most favorable change in free energy binding energy through the twenty years (-66.73 kcal $mol^{-1}$). This may be due to continuous pressure by the original 1968 antibody on new mutated viruses, in addition to their own HA specific antibodies. That is, if any of the earlier 1968 HA viruses were still circulating in the human population, antibodies against it would also be present; new viruses would therefore try to mutate to prevent strong binding by these antibodies as well as their most immediate-specific antibody. Interestingly, despite the overall positive trend of decrease in energy, the sequential mutations showed a significant fluctuation in binding energy. This might be additional proof that the mutations are random but self-corrective in nature. Sites where amino acid changes can alter antigenicity are so limited that in some strains key amino acids in these sites changed for the second time as evolution proceeded while some reverted back. Presumably, at any stage in the evolution of a viral subtype, several different strains may be circulating simultaneously and the combination of changes that are most favorable to the virus would be ones that are preserved during evolution [17, 31]. Successful strains dominate and thus become progenitors for future generations; these are ones that have HAs that don't bind too tightly to currently available antibodies.

A search of available literature did not result in any articles that indicated any general preference of amino acid by a larger or different side chain for mutations. This was the case in this study. However, from a quick look at the types of mutations of the

HAs analyzed, it seems there were was a general conservation of mutations in terms of polarity i.e. a polar amino acid was substituted with another polar amino acid. In addition it has been suggested that oligosaccharide attachment also prevents antibody recognition [93]. This could be due to the inability of carbohydrate covered protein side chain surfaces to induce antibodies, as the carbohydrates side chains are synthesized by cellular enzymes and are thus "antigenically self". They thus mask regions of proteins from recognition by antibody. It should be noted that the lack of information on the crystalline structure of oligosaccharide attachments in HA herein studied, implies that the analysis of these attachments were not included as part of this study. This is despite the fact that since the beginning of the Hong Kong pandemic period in 1968, a study of the recent changes in influenza virus showed an increase in the number of oligosaccharide attachment sites (especially on the membrane distal domain of HA). While the Hong Kong HA monomer in 1968 had a total of three attachment sites (at position 81, 165, 285), this number had increased to eight by 2005 (63, 122, 126, 133, 144, 165, 246 and 285) [23, 94]. A study of the role of oligosaccharide attachment in reduction in binding affinity would be beneficial in better predicting the trend in amino acid substitution during HA antigenicity.

Detailed analyses of the binding energetics of influenza HA to the 1968 antibody allows better understanding of the contribution of different energy terms to the binding free energy. The association of the antibody and HA is driven by highly favorable van der Waals interactions ($\Delta E_{VDW}$), the most important and consistent complex formation favoring term (Figure 15). This being the case, it is surprising to note that its contribution is constant and does not decrease with time to allow the virus to more easily escape

antibody binding. In all seven HA-AB complexes investigated, the van der Waals interactions for HA binding were between -80.6 and -94.6 kcal $mol^{-1}$. In line with the results of this work, the non-covalent association has been found to be driven by favorable van der Waals energies in several MM-PBSA studies [75] .

In contrast to van der Waals energies, electrostatic energies contributed unequally to the binding of the HA virus strains to the 1968 antibody. That the unfavorable electrostatic solvation is compensated, but not fully, by favorable electrostatic protein–ligand interactions has been observed in several earlier ligand-binding studies and seems to be a general phenomenon [75, 76]. In agreement with earlier computational ligand binding studies, the solvation electrostatic energies ($\Delta\Delta G_{elec,\ PB/\ GB}$) are mostly unfavorable for binding, but the degree at which the energy disfavors binding, decreases with year progression. For example by 1999 the solvation electrostatic energy favors binding compared to 1968 when it disfavors binding. The exact opposite effect is seen with solute electrostatic energies ($\Delta\Delta E_{ELE}$). While in the 1968 and 1969 HA the energies favor binding, by 1999 the energies disfavor binding. Thus, electrostatic interactions determine the specificity in the binding of influenza HA to antibody. Solvent effects are most important when the binding between HA and antibodies is less strong. Looking back at the total charges of the modeled HAs (Table 1), mutations seem to favor amino acids that make the HA more positively charged (3 for 68 vs. +6 for 99). This could account for the $\Delta E_{ELE,}$ electrostatic molecular mechanical energy contribution calculated from the Coulomb potential becoming more positive with year progression (thus disfavoring complex formation). Less favorable solute energies due to the preference of more positively charged amino acids during mutation could hint to a strategy used by influenza

viruses to escape antibody binding. Analyses of more viruses should provide more evidence of this trend [95]. Similar to other studies [74, 77], both internal energy ($\Delta E_{INT}$) and non-electrostatic energies contribution to solvation ($\Delta\Delta G_{np}$) are very small and contribute minimally to the complex formations.

The MM-GBSA calculations of the study seem to over estimate the relative free binding energy of influenza HA to antibody when compared to the limited available experimental values (12 kcal $mol^{-1}$ [3]) and other free energy computational studies (36 vs. 66 kcal $mol^{-1}$ [96]). However, in view of the fact that the accuracy of absolute binding free energy calculations depends on a delicate balance of different energetic and entropic contributions, the calculated binding free energy can still be considered close to the experimentally determined one. In addition, for the purpose of this study, we were more interested in the trend of binding energies (qualitative) than the exact values (quantitative). Therefore, if we assume that all values were over estimated by the same degree, the conclusions that the trend shows a decrease in negative binding energy with increase in mutations, remains valid. It should also be noted that the study involved stimulation of the HA monomer and not the whole homotrimmer, as it is found in nature. This could have also contributed to the difference in results.

Calculations of the binding free energies of the HA–AB for snapshots extracted from the single complex trajectory versus three trajectories calculations might have contributed to the difference in values from those that have previously described in experiments. The method was chosen because it is less time consuming than using snapshots from three separate trajectories and potentially requires less sampling, because all of the intramolecular energies cancel when calculating the association energy. Related

studies applying MM-PB(GB)SA for the prediction of absolute binding affinities have shown good agreement between values calculated for snapshots extracted either from the complex trajectory or from all three trajectories. However, Gohlke et al [74, 97] showed that the single trajectories approach is only completely successful, if no conformational and dynamical changes upon complex formation occur in the binding partners. In the case of the influenza HA although there has been some evidence of conformational change upon binding [48, 98], this has not been proven. RMSD and B-Factor plots made of the 1968 HA and 1985 HA (Figure 14) suggest that although the antibody is more flexible than the HA during the simulation, the HA is undergoing a more significant conformation shift and is taking longer to equilibrate. If conformational change does take place, it would account for some of the disparity between the calculated binding free energies and the experimental values available.

Another minor complication that resulted was that despite a similar qualitative pattern in binding energy data being realized in both GB and PB methods (which as mentioned earlier is what was most important in this study), a difference in quantitative data was obtained by the GB and PB method. The difference in values of binding energies by the PB or GB method results from a difference in calculation of the electrostatic energies by solving the respective equations, as shown in Table 3. The difference could be due to choice of parameters. Upon HA-AB complex formation, polar and charged residues become buried in the binding interface. Modeling of the desolvation penalty and correctly estimating screened electrostatic interactions between both proteins is therefore crucial for the accurate estimation of binding affinity. Free energy contributions can be calculated by solving the Poisson-Boltzmann equation using radii

from the PARSE parameter set (PB-linearized or PB-non-linearized) Bondi radii or by solving the General Born equation by applying generalized Born models of Jayaram et al. (MGB/GB $^{JSB}$) [99], Onufriev et al. (GB$^{OBC}$) [100], and Tsui et al (GB$^{HCT}$) [101]. The different models using the different parameter sets produce different results as shown by Golkhe and Case [74] who tested the internal consistency and the model dependence influence of different continuum solvation models on the absolute binding free energy of Ras-Raf . They found that calculated absolute binding free energies strongly depend on the applied solvation model and were within a range of 6.2 kcal mol$^{-1}$ (from PB equation using an linearized PARSE parameter) to -49.4 kcal mol$^{-1}$ (from GB equation using Tsui et al (GB$^{HCT}$) solvation model). The results obtained using different GB models in connection with a solvent-accessible surface area-dependent term (to estimate nonpolar solvation energies) overestimated the binding affinity of Raf towards Ras, whereas PB calculations yielded binding free energies that were either too positive or too negative, depending on the atomic radii applied. Studies have shown that using the GB$^{OBC}$ would result in the results that are most similar to PB results. This is not unexpected, because GB$^{OBC}$ has been modified to give better agreement between PB and GB results for RAS-RAF binding. It is thought that PB gives values that are closest to experimental values. No studies were found that showed the same agreement in values between the PB and GB method when using the GB$^{OBC}$ model in other protein-antibody systems. In this study, the GB$^{HCT}$ model was chosen for the GB calculation, as it was the only model that worked with the current parameter set and system set up. This could be due to the AMBER version used as some of the GB models require later versions of AMBER.

Further testing and parameterizations across large sets of systems with different properties will be necessary to probe the transferability of the protocols and further improve parameters for influenza. It will be interesting to evaluate the influence of recent developments in the field of polarizable force fields and continuum solvent descriptions on the outcome of these calculations.

Even though the basic idea is simple, the art of MD was challenging in practice as the results of a MD simulations can only be as good as the governing force field. The functional forms of force fields employed in the molecular mechanics were a compromise between accuracy and computational efficiency. For example, that the calculated binding free energies were larger than those obtained by experiment may have also been caused by the short time scale of the MD simulations. Also due to the computational demand, the calculations were reported for 60 snapshots from the set of 500 extracted from trajectories. Studies done using only a subset of snapshots did not lead to a significant change of the calculated binding free energy compared to the one obtained for the total set; but there was an increase in standard deviation. These simplifications meant that the MD simulations of the trajectories needed for energy analyses was by today's standards a computationally inexpensive tasks, making screening of the multiple influenza complexes feasible in a limited amount of time.

Other various techniques that could be used to predict relative binding free energies of protein-protein association include free energy perturbation (FEP) [95] and linear interaction energy (LIE). FEP and other similar computer simulation techniques, such as thermodynamic integration, offer a rigorous statistical mechanical way to calculate free energies of binding. However, for the purpose of this study, they were not

conducive as they also suffer from long computational times and technical difficulties associated with creation and annihilation of atoms and have typically been used for single mutation studies. Also for an LIE calculation experimental binding data is needed to fit against in order to do the calculation, which is rarely available [102]. Compared to the more traditional Free Energy Perturbation (FEP) simulations, the MM–PBSA method is a slightly less accurate but a considerably faster approach. However, it is important to note that although the FEP methods can be applied reliably only to relatively small mutations, it in favorable cases provides free energies at chemical accuracy (i.e. within 1 kcal $mol^{-1}$ from experiment). However, this accuracy would not have been beneficial in the case of this study since the computing time for similar study using FEP would be too expensive and this study involved doing multiple mutations.

The success in calculating and summarizing the relative binding free energies of the HAs indicates that the MM-PBSA method is a fairly reliable tool to investigate virus-antibody interaction that is not too computer time expensive. The results show that antigenic drift of influenza viruses caused by gradual mutations to the HA gene continually produce immunologically distinct strains that bind less tightly to the specific antibody. Since the emergence of such new drift variants underlines the importance of influenza virus surveillance, the effectiveness of influenza vaccinations may be greatly reduced when there is a significant antigenic mismatch between vaccine strains and circulating viruses [103]. Although our data does provide a bit more insight on the trend of mutations, neither antigenic nor energetic analysis alone can give sufficient information on the route of evolution of influenza viruses. Therefore, additional antigenic as well as

genetic characterization of currently circulating influenza viruses is required for determining the most appropriate composition of new influenza vaccines.

# REFERENCES


1.Wiley, D. C.; Wilson, I. A.; Skehel, J. J., Structural Identification of the Antibody-Binding Sites of Hong Kong Influenza Hemagglutinin and Their Involvement in Antigenic Variation. *Nature* **1981,** *289* (5796), 373-8.

2.Tulip, W. R.; Varghese, J. N.; Webster, R. G.; Laver, W. G.; Colman, P. M., Crystal Structures of Two Mutant Neuraminidase-Antibody Complexes with Amino Acid Substitutions in the Interface. *J Mol Biol* **1992,** *227* (1), 149-59.

3.Fleury, D.; Wharton, S. A.; Skehel, J. J.; Knossow, M.; Bizebard, T., Antigen Distortion Allows Influenza Virus to Escape Neutralization. *Nat Struct Biol* **1998,** *5* (2), 119-23.

4.Dan, H. B., Rational Design of Gene-Based Vaccines. *J Pathol* **2006,** *208* (2), 283-9.

5.Verlinde, C. L.; Hol, W. G., Structure-Based Drug Design: Progress, Results and Challenges. *Structure* **1994,** *2* (7), 577 - 87.

6.Nicholson, K. G., Clinical Features of Influenza. *Semin Respir Infect* **1992,** *7* (1), 26-37.

7.Klimov, A.; Simonsen, L.; Fukuda, K.; Cox, N., Surveillance and Impact of Influenza in the United States. *Vaccine* **1999,** *17 Suppl 1*, S42-6.

8.Turkulov, V.; Madle-Samardzija, N., Influenza--Always Present among Us. *Med Pregl* **2000,** *53* (3-4), 154-8.

9.Cheung, T. K.; Poon, L. L., Biology of Influenza A Virus. *Ann N Y Acad Sci* **2007,** *1102*, 1-25.

10.Maeda, Y.; Horimoto, T.; Kawaoka, Y., Classification and Genome Structure of Influenza Virus. *Nippon Rinsho* **2003,** *61* (11), 1886-91.

11.Sidorenko, Y.; Reichl, U., Structured Model of Influenza Virus Replication in MDCK Cells. *Biotechnol Bioeng* **2004,** *88* (1), 1-14.

12.Skehel, J. J.; Wiley, D. C., Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin. *Annu Rev Biochem* **2000,** *69*, 531-69.

13.Zambon, M. C., Epidemiology and Pathogenesis of Influenza. *J Antimicrob Chem other* **1999,** *44 Suppl B*, 3-9.

14.Ohuchi, M.; Asaoka, N.; Sakai, T.; Ohuchi, R., Roles of Neuraminidase in the Initial Stage of Influenza Virus Infection. *Microbes Infect* **2006,** *8* (5), 1287-93.

15.WHO, A Revision of the System of Nomenclature for Influenza Viruses: A Revision of the System of Nomenclature for Influenza Viruses: A WHO Memorandum. World Health Organization: 1980; Vol. 58, pp 585-91.

16.Laver, W. G.; Air, G. M.; Webster, R. G.; Gerhard, W.; Ward, C. W.; Dopheide, T. A., Antigenic Drift in Type A Influenza Virus: Sequence Differences in the Hemagglutinin of Hong Kong (H3N2) Variants Selected with Monoclonal Hybridoma Antibodies. *Virology* **1979,** *98* (1), 226-37.

17.Both, G. W.; Sleigh, M. J.; Cox, N. J.; Kendal, A. P., Antigenic Drift in Influenza Virus H3 Hemagglutinin from 1968 to 1980: Multiple Evolutionary Pathways and Sequential Amino Acid Changes at Key Antigenic Sites. *J Virol* **1983,** *48* (1), 52-60.

18.Raymond, F. L.; Caton, A. J.; Cox, N. J.; Kendal, A. P.; Brownlee, G. G., The Antigenicity and Evolution of Influenza H1 Hemagglutinin, from 1950-1957 and 1977-1983: Two Pathways from One Gene. *Virology* **1986,** *148* (2), 275-87.

19.Air, G. M., Sequence Relationships among the Hemagglutinin Genes of 12 Subtypes of Influenza A Virus. *Proc Natl Acad Sci U S A* **1981,** *78* (12), 7639-43.

20.Colman, P. M.; Varghese, J. N.; Laver, W. G., Structure of the Catalytic and Antigenic Sites in Influenza Virus Neuraminidase. *Nature* **1983,** *303* (5912), 41-4.

21.Nobusawa, E.; Aoyama, T.; Kato, H.; Suzuki, Y.; Tateno, Y.; Nakajima, K., Comparison of Complete Amino Acid Sequences and Receptor-Binding Properties among 13 Serotypes of Hemagglutinins of Influenza A Viruses. *Virology* **1991,** *182* (2), 475-85.

22.Laver, W. G.; Air, G. M.; Webster, R. G., Mechanism of Antigenic Drift in Influenza Virus. Amino Acid Sequence Changes in an Antigenically Active Region of Hong Kong (H3N2) Influenza Virus Hemagglutinin. *J Mol Biol* **1981,** *145* (2), 339-61.

23.Lin, Y. P.; Gregory, V.; Bennett, M.; Hay, A., Recent Changes among Human Influenza Viruses. *Virus Res* **2004,** *103* (1-2), 47-52.

24.Hayden, F.; Croisier, A., Transmission of Avian Influenza Viruses to and between Humans. *J Infect Dis* **2005,** *192* (8), 1311-4.

25.Nelson, M. I.; Holmes, E. C., The Evolution of Epidemic Influenza. *Nat Rev Genet* **2007,** *8* (3), 196-205.

26.Cox, R. J.; Brokstad, K. A.; Ogra, P., Influenza Virus: Immunity and Vaccination Strategies. Comparison of the Immune Response to Inactivated and Live, Attenuated Influenza Vaccines. *Scand J Immunol* **2004,** *59* (1), 1-15.

27.Wiley, D. C.; Skehel, J. J.; Waterfield, M., Evidence from Studies with a Cross-Linking Reagent That the Hemagglutinin of Influenza Virus Is a Trimer. *Virology* **1977,** *79* (2), 446-8.

28.Wilson, I. A.; Skehel, J. J.; Wiley, D. C., Structure of the Hemagglutinin Membrane Glycoprotein of Influenza Virus at 3Å Resolution. *Nature* **1981,** *289* (5796), 366-73.

29.Caton, A. J.; Brownlee, G. G.; Yewdell, J. W.; Gerhard, W., The Antigenic Structure of the Influenza Virus A/PR/8/34 Hemagglutinin (H1 Subtype). *Cell* **1982,** *31* (2 Pt 1), 417-27.

30.Gerhard, W.; Yewdell, J.; Frankel, M. E.; Webster, R., Antigenic Structure of Influenza Virus Hemagglutinin Defined by Hybridoma Antibodies. *Nature* **1981,** *290* (5808), 713-7.

31.Sleigh, M. J.; Both, G. W.; Underwood, P. A.; Bender, V. J., Antigenic Drift in the Hemagglutinin of the Hong Kong Influenza Subtype: Correlation of Amino Acid Changes with Alterations in Viral Antigenicity. *J Virol* **1981,** *37* (3), 845-53.

32.Sauter, N. K.; Hanson, J. E.; Glick, G. D.; Brown, J. H.; Crowther, R. L.; Park, S. J.; Skehel, J. J.; Wiley, D. C., Binding of Influenza Virus Hemagglutinin to Analogs of Its Cell-Surface Receptor, Sialic Acid: Analysis by Proton Nuclear Magnetic Resonance Spectroscopy and X-Ray Crystallography. *Biochemistry* **1992,** *31* (40), 9609-21.

33.Weis, W.; Brown, J. H.; Cusack, S.; Paulson, J. C.; Skehel, J. J.; Wiley, D. C., Structure of the Influenza Virus Hemagglutinin Complexed with Its Receptor, Sialic Acid. *Nature* **1988,** *333* (6172), 426-31.

34.Brownlee, G. G.; Fodor, E., The Predicted Antigenicity of the Hemagglutinin of the 1918 Spanish Influenza Pandemic Suggests an Avian Origin. *Philos Trans R Soc Lond B Biol Sci* **2001,** *356* (1416), 1871-6.

35.Wiley, D. C.; Skehel, J. J., The Structure and Function of the Hemagglutinin Membrane Glycoprotein of Influenza Virus. *Annu Rev Biochem* **1987,** *56*, 365-94.

36.Ha, Y.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C., X-Ray Structures of H5 Avian and H9 Swine Influenza Virus Hemagglutinins Bound to Avian and Human Receptor Analogs. *Proc Natl Acad Sci U S A* **2001,** *98* (20), 11181-6.

37.Russell, R. J.; Stevens, D. J.; Haire, L. F.; Gamblin, S. J.; Skehel, J. J., Avian and Human Receptor Binding by Hemagglutinins of Influenza A Viruses. *Glycoconj J* **2006,** *23* (1-2), 85-92.

38.Ha, Y.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C., H5 Avian and H9 Swine Influenza Virus Hemagglutinin Structures: Possible Origin of Influenza Subtypes. *EMBO J* **2002,** *21* (5), 865-75.

39.Beigel, J. H.; Farrar, J.; Han, A. M.; Hayden, F. G.; Hyer, R.; de Jong, M. D.; Lochindarat, S.; Nguyen, T. K.; Nguyen, T. H.; Tran, T. H.; Nicoll, A.; Touch, S.; Yuen, K. Y., Avian Influenza A (H5N1) Infection in Humans. *N Engl J Med* **2005,** *353* (13), 1374-85.

40.Thompson, C. I.; Barclay, W. S.; Zambon, M. C.; Pickles, R. J., Infection of Human Airway Epithelium by Human and Avian Strains of Influenza A Virus. *J Virol* **2006,** *80* (16), 8060-8.

41.Nicholls, J. M.; Bourne, A. J.; Chen, H.; Guan, Y.; Peiris, J. S., Sialic Acid Receptor Detection in the Human Respiratory Tract: Evidence for Widespread Distribution of Potential Binding Sites for Human and Avian Influenza Viruses. *Respir Res* **2007,** *8*, 73.

42.Guo, C. T.; Takahashi, N.; Yagi, H.; Kato, K.; Takahashi, T.; Yi, S. Q.; Chen, Y.; Ito, T.; Otsuki, K.; Kida, H.; Kawaoka, Y.; Hidari, K. I.; Miyamoto, D.; Suzuki, T.; Suzuki, Y., The Quail and Chicken Intestine Have Sialyl-Galactose Sugar Chains Responsible for the Binding of Influenza A Viruses to Human Type Receptors. *Glycobiology* **2007,** *17* (7), 713-24.

43.NCBI Scheme of Influenza A Virus Replication. http://www.ncbi.nlm.nih.gov/genomes/GenomesHome.cgi?taxid=10239&hopt=scheme (accessed 5 July 2009).

44.Knossow, M.; Gaudier, M.; Douglas, A.; Barrere, B.; Bizebard, T.; Barbey, C.; Gigant, B.; Skehel, J. J., Mechanism of Neutralization of Influenza Virus Infectivity by Antibodies. *Virology* **2002,** *302* (2), 294-8.

45.Knossow, M.; Skehel, J. J., Variation and Infectivity Neutralization in Influenza. *Immunology* **2006,** *119* (1), 1-7.

46.Fleury, D.; Barrere, B.; Bizebard, T.; Daniels, R. S.; Skehel, J. J.; Knossow, M., A Complex of Influenza Hemagglutinin with a Neutralizing Antibody That Binds Outside the Virus Receptor Binding Site. *Nat Struct Biol* **1999,** *6* (6), 530-4.

47.Barbey-Martin, C.; Gigant, B.; Bizebard, T.; Calder, L. J.; Wharton, S. A.; Skehel, J. J.; Knossow, M., An Antibody That Prevents the Hemagglutinin Low pH Fusogenic Transition. *Virology* **2002,** *294* (1), 70-4.

48.Bizebard, T.; Gigant, B.; Rigolet, P.; Rasmussen, B.; Diat, O.; Bosecke, P.; Wharton, S. A.; Skehel, J. J.; Knossow, M., Structure of Influenza Virus Hemagglutinin Complexed with a Neutralizing Antibody. *Nature* **1995,** *376* (6535), 92-4.

49.Amit, A. G.; Mariuzza, R. A.; Phillips, S. E.; Poljak, R. J., Three-Dimensional Structure of an Antigen-Antibody Complex at 2.8Å Resolution. *Science* **1986,** *233* (4765), 747-53.

50.Dodson, G. G.; Lane, D. P.; Verma, C. S., Molecular Simulations of Protein Dynamics: New Windows on Mechanisms in Biology. *EMBO Rep* **2008,** *9* (2), 144-50.

51.Van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B., Biomolecular Modeling: Goals, Problems, Perspectives. *Angew Chem Int Ed Engl* **2006,** *45* (25), 4064-92.

52.Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A., Energy Parameters in Polypeptides. Vii. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids. *J Phys Chem* **1975,** *79* (22), 2361-81.

53.Momany, F. A.; Carruthers, L. M.; McGuire, R. F.; Scheraga, H. A., Intermolecular Potentials from Crystal Data. Iii. Determination of Empirical Potentials and Application to the Packing Configurations and Lattice Energies in Crystals of Hydrocarbons, Carboxylic Acids, Amines, and Amides. *J Phys Chem* **1974,** *78* (16), 1595-620.

54.Hagler, A. T.; Lifson, S.; Dauber, P., Consistent Force Field Studies of Intermolecular Forces in Hydrogen-Bonded Crystals. 2. A Benchmark for the Objective Comparison of Alternative Force Fields. *J Am Chem Soc* **1979,** *101* (18), 5122-30.

55.Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, I., T.E; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. A., AMBER, a Package of Computer Programs for Applying Molecular Mechanics, Normal Mode Analysis, Molecular Dynamics and Free Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Comp Phys Commun* **1995,** *91*, 1-41.

56.Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* **1995,** *117* (19), 5179-97.

57.Gillespie, R. J.; Bayles, D.; Platts, J.; Heard, G. L.; Bader, R. F. W., The Lennard-Jones Function & NBSP; a Quantitative Description of the Spatial Correlation of Electrons as Determined by the Exclusion Principle. *J Phys Chem A* **1998,** *102* (19), 3407-14.

58.Ullah, N., Screened Coulomb Potential Problem in the Momentum Representation. *Phys Rev A* **1989,** *40* (12), 6831.

59.Becker, P.; Coppens, P., About the Coulomb Potential in Crystals. *Acta Crystallographica Section A* **1990,** *46* (4), 254-8.

60.Fletcher, R.; Reeves, C. M., Function Minimization by Conjugate Gradients. *Comput J* **1964,** *7* (2), 149-54.

61.Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., The Missing Term in Effective Pair Potentials. *J Phys Chem* **1987,** *91* (24), 6269-71.

62.Jorgensen, W. L.; Madura, J. D., Quantum and Statistical Mechanical Studies of Liquids. 25. Solvation and Conformation of Methanol in Water. *J Am Chem Soc* **1983,** *105* (6), 1407-13.

63.Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A., H++: A Server for Estimating pKas and Adding Missing Hydrogens to Macromolecules. *Nucleic Acids Res* **2005** *33*, 368-71.

64.Anandakrishnan, R.; Onufriev, A., Analysis of Basic Clustering Algorithms for Numerical Estimation of Statistical Averages in Biomolecules. *J Comput Biomol* **2008,** *15*, 165-84.

65.Onufriev, A. H++. http://biophysics.cs.vt.edu/H++ (accessed 7th August 2007).

66.Darden, T. Y., Darrin; Pedersen, Lee, Particle Mesh Ewald: An N . Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **1993,** *98* (12), 10089-92.

67.Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J Comput Phys* **1977,** *23* (3), 327-41.

68.Levine, I. N., *Physical Chemistry*. McGraw Hill Brooklyn College, City of New York, 1978; p 847.

69.Müller, I., *A History of Thermodynamics - the Doctrine of Energy and Entropy*. Springer: 2007.

70.Jayaram, B.; McConnell, K. J.; Surjit, B. D.; Beveridge, D. L., Free Energy Analysis of Protein-DNA Binding: The EcoRI Endonuclease-DNA Complex. *J Comput Phys* **1999,** *151* (1), 333-57.

71.Jayaram, B.; McConnell, K. J.; Dixit, S. B.; Das, A.; Beveridge, D. L., Free-Energy Component Analysis of 40 Protein-DNA Complexes: A Consensus View on the Thermodynamics of Binding at the Molecular Level. *J Comput Chem* **2002** *23*, 1-14.

72.Kombo, D. C.; Jayaram, B.; McConnell, K. J.; Beveridge, D. L., Calculation of the Affinity of the Lambda Repressor-Operator Complex Based on Free Energy Component Analysis. *Mol Simulat* **2002,** *28* (1), 187-211.

73.Ajay; Murcko, M. A., Computational Methods to Predict Binding Free Energy in Ligand-Receptor Complexes. *J Med Chem* **1995,** *38* (26), 4953-67.

74.Gohlke, H.; Case, D. A., Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J Comput Chem* **2004,** *25* (2), 238-50.

75.Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., 3rd, Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc Chem Res* **2000,** *33* (12), 889-97.

76.Huo, S.; Massova, I.; Kollman, P. A., Computational Alanine Scanning of the 1:1 Human Growth Hormone-Receptor Complex. *J Comput Chem* **2002,** *23* (1), 15-27.

77.Swanson, J. M.; Henchman, R. H.; McCammon, J. A., Revisiting Free Energy Calculations: A Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys J* **2004,** *86* (1 Pt 1), 67-74.

78.Srinivasan, J.; Miller, J.; Kollman, P. A.; Case, D. A., Continuum Solvent Studies of the Stability of RNA Hairpin Loops and Helices. *J Biomol Struct Dyn* **1998,** *16* (3), 671-82.

79.Protein Data Bank (PDB Entry Code ID, 1EO8) http://www.rcsb.org/ (accessed 8th August 2007).

80.Influenza Virus Resource. http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html (accessed 7th August 2007).

81.Altschul, S. F.; Thomas, L. M.; Alejandro, A. S.; Jinghui, Z.; Zheng, Z.; Webb, M.; David, J. L., Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research* **1997,** *25*, 3389-402.

82.Guex, N.; Peitsch, M. C., Swiss-Model and the Swiss-PDBviewer: An Environment for Comparative Protein Modeling. *Electrophoresis* **1997,** *18* 2714-23.

83.Sayle, R.; Milner-White, E. J., Rasmol: Biomolecular Graphics for All. *TIBS* **1995,** *20* (9), 374.

84.Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular Dynamics with Coupling to an External Bath. *J Chem Phys* **1984,** *81* (8), 3684-90.

85.Case, D. A.; Pearlman, D. A.; Caldwell, J. C.; Cheatham, T. E. I.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 7*, University of California: San Francisco, 2002.

86.Weiser, J.; Shenkin, P. S.; Still, W. C., Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J Comput Chem* **1999,** *20* (2), 217-30.

87.Rocchia, W.; Alexov, E.; Honig.E., Extending the Applicability of the Nonlinear Poisson-Boltzmann Equation: Multiple Dielectric Constants and Multivalent Ions. *J Phys Chem B* **2001,** *105* (28), 6507-14.

88.Honig, B.; Nicholls, A., Classical Electrostatics in Biology and Chemistry. *Science* **1995,** *268* (5214 ), 1144-9.

89.Nicholls, A.; Honig, B. H., A Rapid Finite Difference Algorithm, Utilizing Successive over-Relaxation to Solve the Poisson-Boltzmann Equation *J Comput Chem* **1991,** *12*, 435-45.

90.Honig, B. Delphi: A Finite Difference Poisson-Boltzmann Solver. http://wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:DelPhi. (accessed 4th April ).

91.Stephen, F.; Altschul, J. C.; E. Gertz, M.; Agarwala, R.; Morgulis, A.; Alejandro, A.; Schaffer A.; Yi-Kuo Y., Protein Database Searches Using Compositionally Adjusted Substitution Matrices. *FEBS J* **2005,** *272*, 5101-9.

92.Knossow, M.; Daniels, R. S.; Douglas, A. R.; Skehel, J. J.; Wiley, D. C., Three-Dimensional Structure of an Antigenic Mutant of the Influenza Virus Hemagglutinin. *Nature* **1984,** *311* (5987), 678-80.

93.Abe, Y.; Takashita, E.; Sugawara, K.; Matsuzaki, Y.; Muraki, Y.; Hongo, S., Effect of the Addition of Oligosaccharides on the Biological Activities and Antigenicity of Influenza A/H3N2 Virus Hemagglutinin. *J Virol* **2004,** *78*, 9605–11.

94.Skehel, J. J.; Stevens, D. J.; Daniels, R. S.; Douglas, A. R.; Knossow, M.; Wilson, I. A.; Wiley, D. C., A Carbohydrate Side Chain on Hemagglutinins of Hong Kong Influenza Viruses Inhibits Recognition by a Monoclonal Antibody. . *Proc Natl Acad Sci USA* **1984,** *81*, 1779–83.

95.Zhou, R.; Das, P.; Royyuru, A. K., Single Mutation Induced H3N2 Hemagglutinin Antibody Neutralization: A Free Energy Perturbation Study. *J Phys Chem B* **2008,** *112* (49), 15813-20.

96.Takamatsu, Y.; Sugiyama, A.; Purqon, A.; Nagao, H.; Nishikawa, K., Binding Free Energy Calculation and Structural Analysis for Antigen-Antibody Co. In *AIP Conference Proceedings*, , 2006; Vol. 832(5), pp 566-9.

97.Gohlke, H.; Kiel, C.; Case, D. A., Insights into Protein-Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras-Raf and Ras-RalGDS Complexes. *J Mol Biol* **2003,** *330* (4), 891-913.

98.Rini, J. M.; Schulze-Gahmen, U.; Wilson, I. A., Structural Evidence for Induced Fit as a Mechanism for Antibody-Antigen Recognition. *Science* **1992,** *255* (5047), 959-65.

99.Jayaram, B.; Sprous, D.; Beveridge, D. L., Solvation Free Energy of Biomacromolecules & NBSP; Parameters for a Modified Generalized Born Model Consistent with the Amber Force Field. *J Phys Chem B* **1998,** *102* (47), 9571-6.

100.Onufriev, A.; Bashford, D.; Case, D. A., Modification of the Generalized Born Model Suitable for Macromolecules. *J Phys Chem B* **2000,** *104* (15), 3712-20.

101.Tsui, V.; Case, D. A., Theory and Applications of the Generalized Born Solvation Model in Macromolecular Simulations." Biopolymers. *Nucleic Acid Sciences* **2001,** *56*, 257-91.

102.Aqvist, J.; Marelius, J., The Linear Interaction Energy Method for Predicting Ligand Binding Free Energies. *Comb Chem High Throughput Screen* **2001,** *4* (8), 613-26.

103.Kendal, A.; Cox, N. J., Forecasting the Epidemic Potential of Influenza Virus Variants Based on Their Molecular Properties. . *Vaccine 3* **1985,** *Suppl*, 263-6.